THE UNIVERSITY
of EDINBURGH

Genetics and Conservation of the Southern South American Conifers

Mauricio Cano Niklitschek

Doctor of Philosophy

THE UNIVERSITY
of EDINBURGH

School of Biological Sciences

Institute of Evolutionary Biology

Royal Botanic Garden of Edinburgh

The University of Edinburgh

March 2020

Abstract

Coniferous trees dominate many northern forests, and globally there are around 627 species of conifers, from eight families and 70 genera. Although the largest abundance of conifers is in the great boreal forests of Eurasia and North America, the species diversity increases further south, even though the land area is smaller. The conifer flora of Chile comprises only nine conifer species, but these include 3/8 extant conifer families (Araucariaceae, Cupressaceae, Podocarpaceae) and eight genera, with all of these species and four genera just restricted to southern South America. Chilean conifers are concentrated in the Chilean Temperate Rainforest, one of the world's plants biodiversity hotspots. Although some species of Chilean conifer such as *Araucaria araucana* (Molina) K. Koch (Monkey Puzzle) have been subject to intensive research, most have not. Knowledge gaps in their basic biology represent a limitation in the development of effective conservation strategies. This is a pressing challenge given the extensive threat to plant biodiversity in the region (harvesting, climate change, pathogens, expanding plantation forestry and agriculture, and natural and human-induced fires). To address this challenge, this research project focuses on assessing population connectivity/differentiation in four emblematic endemic conifers from South America each with a restricted area of distribution: *Saxegothaea conspicua* Lindley, *Prumnopitys andina* (Poepp. ex Endl) de Laubenfels, *Podocarpus salignus* D. Don all members of the Podocarpaceae; and *Fitzroya cupressoides* (Molina) Johnston (Cupressaceae). Between seven and ten populations (per species) were included in this investigation, covering the entire natural distribution of each conifer species in Chile. Restriction site-associated DNA markers (RAD-seq) analysis was conducted. RAD-seq was first optimised (*de novo* assembly) to accommodate the large and complex genomes of conifers, before being deployed to assess patterns of population genetic structure. The main finding of population genetic analyses was a similar level of nucleotide diversity and low levels of genetic structure in all four conifer species. The low population structure over relatively large geographical distances was particularly noteworthy. This result is likely due to the extreme longevity of individuals of these species restricting divergence due to genetic drift, despite their currently fragmented ranges. A regeneration survey was also conducted for one of the

conifer species involved in this project (*Pr. andina*), and supplemented with informal observations on regeneration for the other three species. Regeneration was frequent throughout most of the natural distribution of *Pr. andina*. However, a very low number of saplings was observed in all of the populations (most of them had no saplings at all). Informal observations on *S. conspicua*, *P. salignus* and *F. cupressoides* suggest that regeneration is common for all these species (both seedlings and saplings). However, there was some variation and *P. salignus* showed a higher level of regeneration than *S. conspicua* and *F. cupressoides*. A field-based observation of threats was also undertaken for each conifer species, following the methodology proposed by the IUCN which is directed on recording and evaluating the impact of threats *in situ*. The most common threats with the highest impact were associated with land-conversion to exotic plantations and livestock. This work represents one of the very few studies optimising RAD-seq data for conifers and provides a combination of field and laboratory data to support conservation planning for these important and iconic species.

Lay summary

There are about 627 species of conifer worldwide. Although some conifer species are widespread and common, others are rare and threatened. A particularly important group of conifers from a conservation perspective, are the nine species growing in Chile. Some of these conifers are ecologically and culturally important e.g. Monkey Puzzle - the National tree of Chile (*Araucaria araucana*) and Alerce a National Monument (*Fitzroya cupressoides*), and other species although less well known, are important as their global distribution is entirely restricted to Chile (*Podocarpus salignus* and *Prumnopitys andina* - endemics). The conifers of Chile face a series of threats including logging, climate change, pathogens, expanding plantation forestry and agriculture, and natural and human-induced fires. In this thesis, I investigate the biology of four of these species, to support the development of effective strategies to conserve them. The species I studied are *Saxegothaea conspicua*, *Prumnopitys andina*, *Podocarpus salignus* and *Fitzroya cupressoides*. They are all long-lived, and one, *Fitzroya*, is one of the longest-lived species on earth, with some individuals more than 3000 years old.

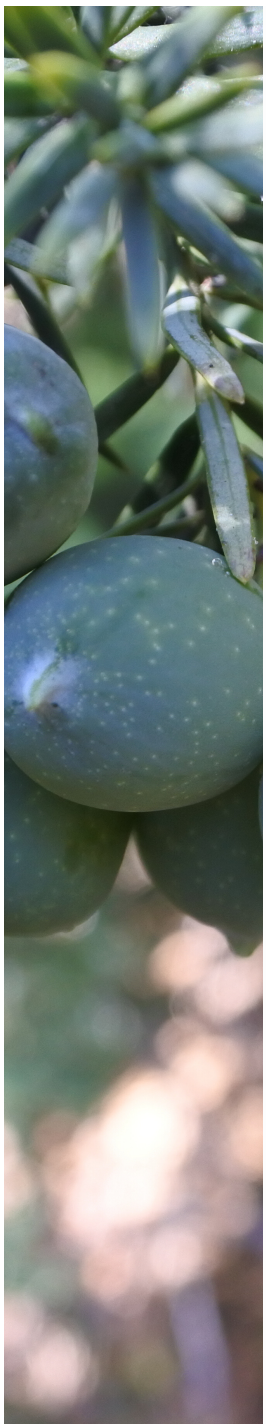
To study these species, I first optimised a method for measuring their genetic variation. I then used this approach to investigate the distribution of genetic variation within each of the species. This showed that although the species have very isolated populations and some small populations, they have little evidence of major genetic diversity loss. I then assessed whether the species showed evidence of healthy regeneration (evidence of seedlings and saplings). Regeneration was frequent throughout most of the natural distribution of *Pr. andina*. However, a very low number of saplings was observed in all of the populations. Less detailed observations on the other three species showed that regeneration is common for all these species (both seedlings and saplings). However, there was some variation and *P. salignus* showed a higher level of regeneration than *S. conspicua* and *F. cupressoides*. I finally combined the genetic work, and the observations on regeneration, with field observations of threats facing each population, and used this combined information to develop a genetic risk assessment to support the conservation of each of these species.

Acknowledgements

First of all, I would like to dedicate this work to three very important people in my life; my godfather Marcelo Cano who, unfortunately, passed away before I could complete this incredible investigation, my best friend Darian Stark, who sadly is not part of this earthly world either, and finally my mother Sandra Niklitschek, who is my friend, my support and an example and inspiration for my life. For all of them, I dedicate this work and my eternal gratitude.

I would like to express my special thanks and gratitude to my main supervisor, Pete Hollingsworth, for his continuous support with my project and in particular for his motivation, interest and extensive knowledge. I was really lucky to have him supporting my work over these four years – I learned so much from him. Many thanks for everything Pete.

Very special thanks to Alex Twyford my other supervisor who helped me in doing this research and on account of whom I came to know about so many new things. I would also like to thank Richard Ennos who was also supporting my research and helped me throughout this process. I am thankful to them. Thanks, of course, to Rachael Brewer, a wonderful person, always doing her best to help on this project.



I also express my deepest gratitude to Markus Ruhsam for his continuous help – thanks for always having time for long discussions and for contributing so much to this project. To Flavia Pezzini, for her patience in expanding my poor knowledge of bioinformatics and even more for her unconditional friendship, thanks a lot Flavia!. To Linda Neaves, for giving me time to discuss my results and for her invaluable contribution to my thesis. To Max Brown for making it easier for me to understand how R works (really thanks) and Wu Huang who also helped me in my bioinformatics work. I would also like to thank Antje Ahrends, Lucia Campos and Philip Thomas who helped me in different aspects of my research and also special thanks to Alex Davey for helping me correct one of my chapters.

I would like to extend my heartiest thanks to all the people who supported and helped me during my expeditions in Chile; Martin Gardner, Peter Baxter, Tom Christian, Ludovica Santilli and Nicolas Lavardero, Fernando Bustos and Reinhard Fitzek, without whom none of this work could have been done. Sincere thanks to all of them.

My eternal gratitude to all the staff of the botanical garden especially to Simon Milne, Ross Irvine, Michelle Hart and Laura Forrest. My thanks also to all landowners and local people from Chile and CONAF who made my work very pleasurable and easy. I also thank to CONICYT for supporting my PhD degree and thanks a lot to Mark Watson and Chris Kettle for they time as examiners and giving me a great VIVA experience.

I also wish to thank some special people in my life without whose support this thesis would have been very difficult to complete. First to Jacqueline Plaza for encouraging me to do this PhD. To my brother and nephews Cristobal, Tomas, Maite and Victoria, to all my cousins and my family in general. To my friends especially to Kai Sedgwick, Andy Griffiths, Sina Marti, Berta, Tas Kartas and Roberto Godoy, the latter for giving me the basis for my professional life. Finally my everlasting thanks to Laura Espona, for all her unconditional support, for her love and friendship. Thank you so much, Laura.

Thanks again to all of you!

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

1.1.1 Conifer conservation

Biodiversity is critical for the maintenance of ecosystem goods and services that benefit humanity, for the functioning of natural ecosystems and planetary bio-geo-chemical cycles, and of intrinsic importance and value as a product of millions of years of evolutionary history. However, human-induced changes to the natural environment are resulting in profound and unprecedented impacts on biodiversity (Poynton et al., 2007). According to the first global assessment report on biodiversity, about 47% of natural ecosystems have declined (with respect to their earliest estimated states), and about 25% of the total species worldwide are already threatened with extinction (Bongaarts, 2019).

Plants are a particularly important element of biodiversity and form the foundation of many of the world's terrestrial habitats, and many plant species are threatened. For instance, conifers are one of the terrestrial plant groups with the highest level of threat. According to recent assessments of the conservation status of conifers using the International Union for Conservation of Nature (IUCN) red-listing framework, about 35% of conifer species are of conservation concern (Bongaarts, 2019), and the most common threats to these species are associated with habitat conversion and degradation, over-exploitation and habitat fragmentation (Newbold et al., 2015). Globally there are only around 627 species of conifers, from eight families and 70 genera, and many of these species are ecologically important (e.g. forming structural components of habitats, influencing local hydrology, capturing carbon, and regulating temperature) (Farjon, 2018). Many conifer species are also of direct value to humans, either as a direct source of products (e.g. timber species such as *Pinus* or sources of medicines such as *Taxus*), or from a perspective of cultural value, and many conifer species are closely associated with culture and traditions of indigenous communities around the world (Turner, 1988).

One particularly important assemblage of conifers is the conifer species of Chile, which includes a set of charismatic species, with distinctive characteristics and many of the species have a discontinuous distribution. An important factor in driving their diversity and distributions is the extraordinary topography of the country, which combines

the Andes range bordering the eastern side of the country, the Central valley providing a distinct low elevation habitat in the middle of the country, and the mountains of the Coastal range at the west of the country bordering the Pacific Ocean.

The Chilean conifers include 3/8 extant conifer families (Araucariaceae, Cupressaceae, Podocarpaceae) with eight genera and nine species (Figure 1.1, 1.2 and 1.3), and with some of these species of a great cultural value for the country; e.g. *Araucaria araucana* (Molina) K. Koch – the monkey puzzle tree. This conifer is an astonishing species with a diameter of up to 2-3 m and 50 m height (Donoso Zegers, 2006). It is also culturally important in the country, usually growing within or near indigenous lands and considered the national tree of Chile, and regarded as a living fossil due to its ancient history (Donoso Zegers, 2006). *Fitzroya cupressoides* (Molina) Johnston is another iconic Chilean conifer, considered a Natural Monument in Chile. This species is probably one of the most long-lived trees in the world (ca. 3500 years) (Lara and Villalba, 1993) and one of the largest trees in South America (up to 5m diameter and 60m height). Other Chilean conifer species include *Austrocedrus chilensis* (D. Don) Pic. Serm. & Bizzarri and *Pilgerodendron uviferum* (D. Don) Florin both from the Cupressaceae family, and five members of the Podocarpaceae family (Table 1.1) with two of them, endemic to Chile; *Prumnopitys andina* (Poepp. ex Endl) de Laubenfels and *Podocarpus salignus* D. Don.

Chilean conifer species face several pressures and have suffered a series of emerging threats over the last decades. Indeed, according to the IUCN, only one species is classed as least concern (*Lepidothamnus fonkii* Phil, Table 1.1). Some of the most common emerging threats to the Chilean conifers include; conversion of land to exotic plantations (*Pinus* and *Eucalyptus*) and/or agricultural activities (livestock), which has caused a severe forests reduction and fragmentation, along with illegal logging, fire and dam construction (hydroelectric schemes) (Aguayo et al., 2009).

Table 1.1: List of conifer species in Chile, showing the distribution of the species (by country), the conservation category according to the IUCN (EN: Endangered, NT: Near threatened, VU: Vulnerable, LC: Least concern) and an estimation of the maximum longevity of each conifer.

Family	**Genera**	**Species**	**Life form**	IUCN	**Distribution**	**Longevity (years)**
Araucariaceae	*Araucaria*	*Araucaria araucana*	Tree	EN	Chile and Argentina	1500
Cupressaceae	*Austrocedrus*	*Austrocedrus chilensis*	Tree	NT	Chile and Argentina	1000
	Fitzroya	*Fitzroya cupressoides*	Tree	EN	Chile and Argentina	2000-3000
	Pilgerodendron	*Pilgerodendron uvifera*	Tree	VU	Chile and Argentina	800
Podocarpaceae	*Lepidothamnus*	*Lepidothamnus fonkii*	Shrub	LC	Chile and Argentina	unknown
	Podocarpus	*Podocarpus nubigenus*	Tree	NT	Chile and Argentina	400
		Podocarpus salignus	Tree	VU	Chile	350
	Prumnopytis	*Prumnopitys andina*	Tree	VU	Chile	600
	Saxegothaea	*Saxegothaea conspicua*	Tree	NT	Chile and Argentina	800 (even more)

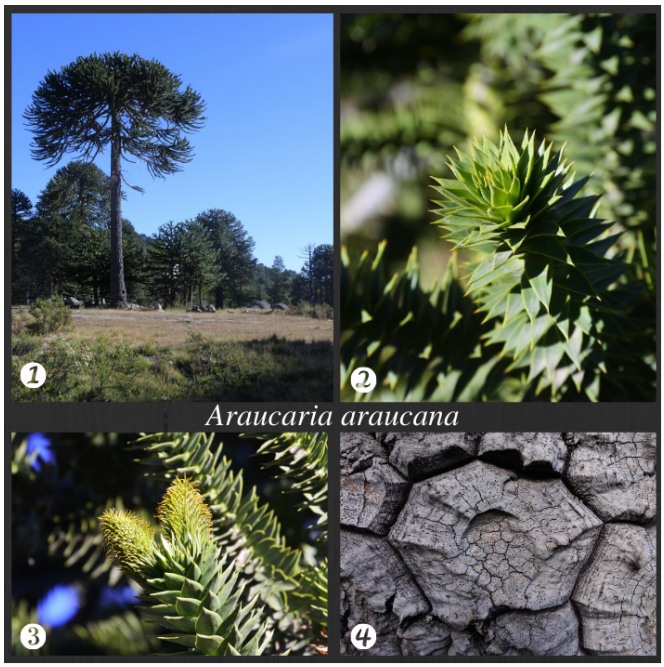


Figure 1.1: *Araucaria araucana* in Chile. 1) adult tree, 2) leaves (needles), 3) male cone, and 4) bark.

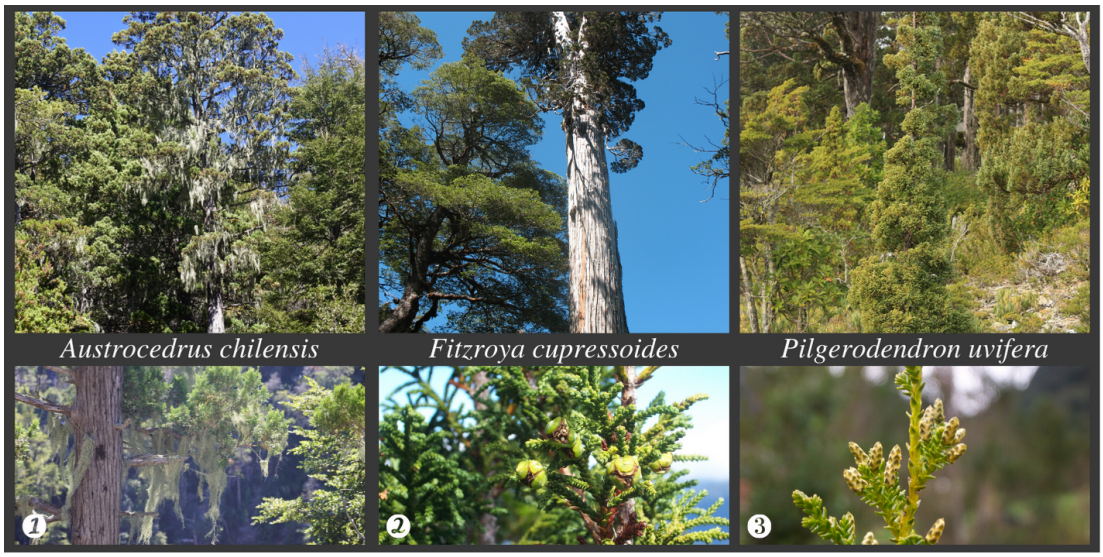


Figure 1.2: Cupressaceae species. Top images show adult trees of each species. Bottom pictures show 1) bark of *Austrocedrus*, 2) female cone of *F. cupressoides* and 3) male cone of *P. uvifera*.

Figure 1.3: Podocarpaceae species in Chile. Top images show adult individuals of each species. Bottom pictures show 1) female cone of *L. fonkii*, 2) fruit of *P. nubigenus*, 3) male cone *P. salignus*, 4) fruit *Pr. andina* and 5) female (in green) and male cone (in brown) of *S. conspicua*.

1.1.2 Population genetics

One aspect of conservation that is receiving increased attention is genetic variation (Laikre et al., 2020). Population genetic studies provide information on the genetic variation within and between populations and can provide insight into ecological and evolutionary processes (i.e. mutation, genetic drift, natural selection, and migration) (Johnston et al., 2019). This is important as the long-term existence of a species is critically influenced by the maintenance of genetic variation within populations (Neto et al., 2014) which ultimately is determined by the size of the population and the levels of gene flow into the population. Maximising genetic variation within populations maximises the likelihood that some individuals have the genetic makeup to survive unpredictable events (Frankham et al., 2004).

Pollen and (to a lesser extent) seed dispersal are the mechanisms for maintaining connectivity between tree populations and maintaining genetic diversity within populations (Burczyk and Koralewski 2005, Hamrick 2004, Liepelt et al. 2002, O'Connell et al. 2007). Populations connected by gene flow should have similar genetic variants, whereas those that are isolated become genetically different (e.g. different alleles present in different populations). Thus studying patterns of genetic variation within and among populations provides insights into patterns of gene flow and connectivity among populations, as well as providing insights into the genetic "health" of populations.

1.1.3 Measuring genetic variation and structure

Choice of genome

Plant DNA is located in three key genomic compartments (nuclear genome, and the organelle genomes – mitochondrial and chloroplast genomes) (Hamilton and Miller 2002, McCauley 1995). In angiosperms, chloroplast and mitochondrial genomes are most often maternally inherited, while in gymnosperms both genomes are predominantly paternally inherited (Corriveau and Coleman 1988, Hamilton and Miller 2002, Mogensen 1996, Reboud and Zeyl 1994), with the notable exception of pines where chloroplast DNA is paternally inherited and the mitochondrial genome is maternally inherited (Wagner, 1992). The organelle genomes can thus provide information about pollen and seed

dispersal, and the ease of working with organelle genomes has led to a large number of studies using them in plant population genetics. However, the vast majority of a plant's DNA is in the nuclear genome, which is biparentally inherited (via pollen and seed) and where recombination generates novel allelic combinations. Thus information from the nuclear genome can provide insights into the mode of reproduction (sexual v/s asexual), breeding behaviour (selfing v/s outcrossing), population structure, gene flow, and also hybridisation and taxon differentiation.

Traditional molecular markers

Over the last few decades, there has been extensive use of genetic markers for understanding population genetic structure and measuring genetic variants. This includes classical molecular markers such as isozymes/allozymes, RAPDs (randomly amplified polymorphic DNA), ISSRs (inter-simple sequence repeats), RFLPs (restriction fragment length polymorphism), AFLPs (amplified fragment length polymorphism) and microsatellites as well as direct amplicon sequencing (Davey et al., 2011)

Although these techniques have proved extremely useful – they are not without limitations (Sunnucks, 2000). RAPDs were notorious for reliability issues (Tyler et al., 1997), and all of the arbitrary fingerprinting techniques (RAPDs, AFLPs, ISSRs) all suffer from the intrinsic problem of indirect access to anonymous DNA polymorphisms making interpretation difficult. RFLPs and amplicon sequencing have been useful for studying the organelle genomes, and Sanger-based amplicon sequencing continues to be popular for studies of variation in plastid DNA, although the amount of variation they detect is limited. Likewise – microsatellites continue to be popular for applications which require very high levels of discriminatory such as paternity analysis and breeding behaviour (e.g. Bacles and Ennos 2008). However, they also suffer from a basic limitation that assaying lots of loci is labour intensive and expensive.

Next Generation Sequencing (NGS)

For more extensive assessment of population genetic structure, next-generation sequencing approaches are desirable as they can provide cost-effective access to large numbers of nuclear markers. Several different NGS platforms are available using different sequencing technologies. However, all of them perform sequencing of millions of small fragments of DNA in parallel, providing high-throughput DNA sequencing (Behjati and Tarpey, 2013). NGS can be used to sequence entire genomes or just a specific area of interest or subsets of the genome (Behjati and Tarpey, 2013). Though NGS platforms provide vast quantities of data, there are also some limitations which are typically associated to error rates (0.1–15%) which are higher and the read lengths generally shorter (35–700) bp for short-read approaches than those of traditional Sanger sequencing platforms (Goodwin et al., 2016). However, regardless of this, NGS approaches are now the mainstream for many population genetic studies. For assessing population structure, genotyping-by-sequencing (GBS) refers to a family of approaches for generating sequence data, typically from large numbers of markers scattered across the nuclear genome. These approaches usually require some form of complexity or reduced-representation methods (Davey et al., 2011) to target sequencing effort to a subset of the genome.

Restriction site associated DNA markers (RAD-Seq)

One reduced representation method applicable to species with limited pre-existing genomic data is RAD-seq. This technique is generally suitable across a wide range of taxa and can be used to answer a wide variety of ecological, evolutionary and conservation-related topics, such as the: genomics of adaptation, inbreeding and genetic diversity, effective population size (Ne), population structure, phylogeography and identification of conservations units (Vendrami et al. 2019, Nunziata and Weisrock 2018, Borrell et al. 2018, Emerson et al. 2010). It provides a high depth of coverage per locus and is a cost-effective method of sequencing hundreds of samples for hundreds or thousands of loci. RAD-seq loci can occur in all areas of the genome (coding and non-coding regions). On the downside, it takes time to optimise, the costs although effective on a per locus base, are far from trivial, and the approach does require a relatively high quality and

quantity of DNA (Davey and Blaxter, 2010).

1.1.4 Population genetics of conifers

General trends

Conifers are typically thought to show low levels of genetic differentiation, even among geographically distant populations (O'Connell et al., 2007). One factor that is postulated as important here is very effective pollen dispersal. However, these observations and generalisations are heavily biased by studies from northern hemispheres boreal conifers like *Pinus*, *Abies*, and *Picea* which frequently occur in large and continuous forest blocks (Dunphy and Hamrick 2007, Hamrick and Murawski 1990, Latouche-Hallé et al. 2004. Nason et al. 1998). Conifers from the southern hemisphere instead often have restricted distributions or at least can be far more restricted in extent than many of the boreal conifer forests. In addition, the southern conifer populations are usually more isolated and fragmented than the northern boreal conifer populations (Farjon, 2018). Despite this, several studies have also concluded low levels of genetic differentiation in many southern hemisphere conifers (e.g. Auler et al. 2002 in *Araucaria angustifolia* (Bertol.) Kuntze; Quiroga and Premoli 2007 in *Podocarpus parlatorei* Pilg).

Population genetics of Chilean conifers

To date, most of the studies documented on population genetics of the Chilean conifers have also shown low levels of genetic differentiation between populations. However, some of the authors have recognised a pattern of genetic differentiation at a regional scale (e.g. between the Andes mountains, Central valley and the Coastal range). For instances, Rafii and Dodd (1998), using epicuticular wax alkane contents, observed genetic differences between Coastal and Andean Chilean populations in *A. araucana*, suggesting ecological adaptations. Bekessy et al. (2002), using RAPD, found a trend of genetic distance with increasing latitude within the Andean *A. araucana* groups, suggesting post-glacial migration route from multiple refugia. Allnutt et al. (1999) using RAPD markers in *F. cupressoides* showed genetic variation between populations, dividing the distribution into three main groups; the north Coastal range, south Coastal

range/Central valley and the Andes. In addition, using isozyme data, Premoli et al. (2000) also revealed that the populations of *F. cupressoides* from the western slope of the Andes were genetically different and more variable than those from elsewhere. However, low levels of genetic variability were observed within populations. Nevertheless, one challenge in interpreting data from these studies is that they were based on a set of molecular markers which all show limitations in terms of understanding genome-wide levels of genetic variation and differentiation; e.g. isozyme markers (conserved loci), or RAPD markers (lack of reproductively).

1.2 Objectives

The underlying motivation behind this thesis has been to generate data on some iconic southern hemisphere conifer species to support their conservation. This reflects the high levels of environmental degradation in Chile and the pressing need for scientific evidence to support conservation interventions, and in particular to integrate information on genetic structure and population processes into the existing knowledge base.

Therefore, this thesis is focused on assessing levels of genetic connectivity/differentiation and patterns of regeneration in four emblematic endemic conifers from South America, each with a restricted and fragmented area of distribution in Chile: *Saxegothaea conspicua* Lindley, *Prumnopitys andina* and *Podocarpus salignus* (all Podocarpaceae species) and *Fitzroya cupressoides* (Cupressaceae).

To achieve this, I have produced multi-locus genotype data from multiple populations covering the entire natural distribution of each species in Chile. RAD sequencing was optimised to accommodate the large and complex genomes of conifers, before being deployed to assess patterns of population genetic structure. The genetic studies have been supported with a formal assessment of patterns of regeneration in one species (*Pr. andina*) and field observations of regeneration in the other three.

This work on regeneration was undertaken in recognition of the fundamental importance of reproduction and recruitment for the maintenance of genetic variation and enabling adaption to environmental change. I also undertook a threat assessment of all four species, to better understand their conservation status.

Thus formally the aims of this thesis are:

1. To develop RAD-seq protocols for four conifer species with large genomes (*S. conspicua*, *Pr. andina*, *P. salignus*, and *F. cupressoides*).
2. To use RAD-seq data to assess population genetic structure in these four species.
3. To quantify levels of regeneration in *Pr. andina*.
4. To review the conservation status of these four conifer species in light of the data produced in this thesis with a particular focus on the conservation of genetic diversity.

1.2.1 Thesis outline

This thesis is divided into five chapters (including this introductory chapter). The next four chapters are presented in the followed order.

- Chapter 2: In this chapter, I focus on the *de novo* assembly of loci from RAD-seq data, with particular attention given to assessing the impact of parameter value changes on the assembly of RAD-seq data in conifer species. Specifically, I assess the optimisation protocol recommended by Paris et al. (2017) and Rochette and Catchen (2017) in four non-model plant species with large genomes and evaluate the sensitivity of parameter settings on assembly and population genetics statistics. This work aims to minimise biases and produce a robust data source for population genetic study.
- Chapter 3: In this chapter, I use the assembled RAD-seq data to investigate the level of genetic diversity and population genetic differentiation in four conifer species. Specifically, I assess (a) general levels of population genetic diversity and population structure, (b) whether any detected differentiation is associated with geographical location (e.g. the Andes vs Coastal range), and (c) whether population genetic variation and levels of differentiation are evenly or unevenly distributed throughout the species ranges.

- Chapter 4: In this chapter, I focus on evaluating the regeneration status of *Pr. andina*, including a characterisation of the *Pr. andina* forests and a description of the main threats found in each population. In an associated Appendix, I also included a full threat assessment for the other three conifer species (*S. conspicua*, *P. salignus* and *F. cupressoides*).

- Chapter 5: The final chapter synthesises the main conclusions from all chapters, including a discussion of the implications of my findings for the conservation of these Chilean conifer species, including a genetic risk assessment, and an outline of plans potential for future work.

Bibliography

Aguayo M, Pauchard A, Azócar G, Parra O. 2009. Cambio del uso del suelo en el centro sur de Chile a fines del siglo XX. Entendiendo la dinámica espacial y temporal del paisaje. *Revista Chilena de Historia Natural*. 82:361–374.

Allnutt TR, Newton AC, Lara A, Premoli A, Armesto JJ, Vergara R, Gardner M. 1999. Genetic variation in *Fitzroya cupressoides* (alerce), a threatened South American conifer. *Molecular Ecology*. 8:975–987.

Auler NMF, Sedrez dos Reis M, Guerra MP, Nodari RO. 2002. The genetics and conservation of *Araucaria angustifolia*: I. Genetic structure and diversity of natural populations by means of non-adaptive variation in the state of Santa Catarina, Brazil. *Genetics and Molecular Biology*. 25:329–338.

Bacles CF, Ennos RA. 2008. Paternity analysis of pollen-mediated gene flow for *Fraxinus excelsior* L. in a chronically fragmented landscape. *Heredity*. 101:368–380.

Behjati S, Tarpey PS. 2013. What is next generation sequencing? *Archives of Disease in Childhood: Education and Practice Edition*. 98:236–238.

Bekessy SA, Allnutt TR, Premoli AC, Lara A, Ennos RA, Burgman MA, Cortes M, Newton AC. 2002. Genetic variation in the Monkey Puzzle tree (*Araucaria araucana* (Molina) K. Koch), detected using RAPD. *Heredity*. 88:243–249.

Bongaarts J. 2019. IPBES, 2019. Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. *Population and Development Review*. 45:680–681.

Borrell JS, Wang N, Nichols RA, Buggs RJ. 2018. Genetic diversity maintained among fragmented populations of a tree undergoing range contraction. *Heredity*. 121:304–318.

Burczyk J, Koralewski TE. 2005. Parentage versus two-generation analyses for estimating pollen-mediated gene flow in plant populations. *Molecular Ecology*. 14:2525–2537.

Corriveau JL, Coleman AW. 1988. Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. *American Journal of Botany*. 75:1443.

Davey JW, Blaxter ML. 2010. RAD-Seq: next-generation population genetics. *Briefings in Functional Genomics*. 9:416–423.

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*. 12:499–510.

Donoso Zegers C. 2006. Las especies arbóreas de los bosques templados de Chile y Argentina : autoecología. Valdivia: Maria Cuneo, primera edition.

Dunphy BK, Hamrick JL. 2007. Estimation of gene flow into fragmented populations of *Bursera simaruba* (Burseraceae) in the dry-forest life zone of Puerto Rico. *American Journal of Botany*. 94:1786–1794.

Emerson KJ, Merz CR, Catchen JM, Hohenlohe PA, Cresko WA, Bradshaw WE, Holzapfel CM. 2010. Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences of the United States of America*. 107:16196–16200.

Farjon A. 2018. The Kew Review Conifers of the World. *Kew Bulletin*. 73:1–16.

Frankham R, Ballou JD, Briscoe DA. 2004. A Primer of Conservation Genetics. Cambridge: Cambridge University Press.

Goodwin S, McPherson JD, McCombie WR. 2016. Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics*. 17:333–351.

Hamilton MB, Miller JR. 2002. Comparing relative rates of pollen and seed gene flow in the island model using nuclear and organelle measures of population structure. *Genetics*. 162:1897–1909.

Hamrick JL. 2004. Response of forest trees to global environmental changes. *Forest Ecology and Management*. 197:323–335.

Hamrick JL, Murawski DA. 1990. The breeding structure of tropical tree populations. *Plant Species Biology*. 5:157–165.

Johnston HR, Keats BJB, Sherman SL. 2019. 12 - Population Genetics. In: Pyeritz RE, Korf BR, Grody WW, editors, Emery and Rimoin's Principles and Practice of Medical Genetics and Genomics, Academic Press, pp. 359–373. Seventh edition.

Laikre L, Hoban S, Bruford MW, et al. (20 co-authors). 2020. Post-2020 goals overlook genetic diversity. *Science*. 367:1085.

Lara A, Villalba R. 1993. A 3620-Year temperature record from *Fitzroya cupressoides* tree rings in southern South America. *Science (New York, N.Y.)*. 260:1104–1106.

Latouche-Hallé C, Ramboer A, Bandou E, Caron H, Kremer A. 2004. Long-distance pollen flow and tolerance to selfing in a neotropical tree species. *Molecular Ecology*. 13:1055–1064.

Liepelt S, Bialozyt R, Ziegenhagen B. 2002. Wind-dispersed pollen mediates postglacial gene flow among refugia. *Proceedings of the National Academy of Sciences of the United States of America*. 99:14590–14594.

McCauley DE. 1995. The use of chloroplast DNA polymorphism in studies of gene flow in plants. *Trends in Ecology & Evolution*. 10:198–202.

Mogensen HL. 1996. The hows and whys of cytoplasmic inheritance in seed plants. *American Journal of Botany*. 83:383–404.

Nason JD, Herre EA, Hamrick JL. 1998. The breeding structure of a tropical keystone plant resource. *Nature*. 391:685–687.

Neto OC, Aguiar AV, Twyford AD, Neaves LE, Toby Pennington R, Lopes AV. 2014. Genetic and ecological outcomes of *Inga vera* subsp. *affinis* (leguminosae) tree plantations in a fragmented tropical landscape. *PLoS ONE*. 9:1–8.

Newbold T, Hudson LN, Hill SL, et al. (41 co-authors). 2015. Global effects of land use on local terrestrial biodiversity. *Nature*. 520:45–50.

Nunziata SO, Weisrock DW. 2018. Estimation of contemporary effective population size and population declines using RAD sequence data. *Heredity*. 120:196–207.

O'Connell LM, Mosseler A, Rajora OP. 2007. Extensive long-distance pollen dispersal in a fragmented landscape maintains genetic diversity in white spruce. *Journal of Heredity*. 98:640–645.

Paris JR, Stevens JR, Catchen JM, Johnston S. 2017. Lost in parameter space: a road map for stacks. *Methods in Ecology and Evolution*. 8:1360–1373.

Poynton JC, Loader SP, Sherratt E, Clarke BT. 2007. Biodiversity hotspots for conservation priorities. *Biodiversity and conservation*. 16:853–858.

Premoli AC, Kitzberger T, Veblen TT. 2000. Isozyme variation and recent biogeographical history of the long-lived conifer *Fitzroya cupressoides*. *Journal of Biogeography*. 27:251–260.

Quiroga MP, Premoli AC. 2007. Genetic patterns in *Podocarpus* parlatorei reveal the long-term persistence of cold-tolerant elements in the southern Yungas. *Journal of Biogeography*. 34:447–455.

Rafii ZA, Dodd RS. 1998. Genetic diversity among coastal and Andean natural populations of *Araucaria araucana* (Molina) K. Koch. *Biochemical Systematics and Ecology*. 26:441–451.

Reboud X, Zeyl C. 1994. Organelle inheritance in plants. *Heredity*. 72:132–140.

Rochette NC, Catchen JM. 2017. Deriving genotypes from RAD-seq short-read data using Stacks. *Nature Protocols*. 12:2640–2659.

Sunnucks P. 2000. Efficient genetic markers for population biology. *Trends in Ecology and Evolution*. 15:199–203.

Turner NJ. 1988. Ethnobotany of coniferous trees in Thompson and Lillooet interior salish of British Columbia. *Economic Botany*. 42:177–194.

Tyler KD, Wang G, Tyler SD, Johnson WM. 1997. Factors affecting reliability and reproducibility of amplification- based DNA fingerprinting of representative bacterial pathogens. *Journal of Clinical Microbiology*. 35:339–346.

Vendrami DL, De Noia M, Telesca L, Handal W, Charrier G, Boudry P, Eberhart-Phillips L, Hoffman JI. 2019. RAD sequencing sheds new light on the genetic structure and local adaptation of European scallops and resolves their demographic histories. *Scientific Reports*. 9:1–13.

Wagner DB. 1992. Nuclear, chloroplast, and mitochondrial DNA polymorphisms as biochemical markers in population genetic analyses of forest trees. *New Forest*. 6:373–390.

Chapter 2

Optimisation and assembly of RAD sequencing data for four conifer species

2.1 Introduction

2.1.1 Next Generation Sequencing

Technological improvements in generating genomic data allow cost-effective investigation of the genomes of a wide variety of non-model organisms, including species with large and complex genomes (e.g. Nowoshilow et al. 2018, Henson et al. 2014, Neale et al. 2014; Birol et al. 2013). High-throughput DNA sequencing is now available on multiple different Next Generation Sequencing (NGS) platforms, using different sequencing technologies, all sequencing millions of DNA fragments in parallel (Behjati and Tarpey, 2013). NGS can be used to sequence entire genomes or specific areas of interest; sequencing subsets of the genome ("reduced representation") has become a very efficient way of studying non-model organisms lacking prior genomic information. One method of reduced representation applicable to species with limited pre-existing genomic data is restriction site-associated DNA sequencing (RAD-seq; Davey et al. 2011). RAD-seq is applicable across a wide range of taxa and can be used to address a variety of ecological, evolutionary and conservation-related topics, such as the genomics of adaptation, inbreeding and genetic diversity, effective population size (Ne), population structure, phylogeography and conservation units (Vendrami et al. 2019, Nunziata and Weisrock 2018, Borrell et al. 2018, Emerson et al. 2010).

2.1.2 Data-processing and software packages

NGS platforms provide vast quantities of data, and thus data processing is a challenge (Díaz-Arce and Rodríguez-Ezpeleta, 2019). Data processing must involve filtering out low-quality reads, assembling or aligning reads, and finally identifying single nucleotide polymorphism (SNPs). Each of these steps requires decisions to be made that depend on the biology of the study system and the nature of the data (Catchen et al., 2013a). Thus, reliable analysis depends on having a statistically rigorous pipeline that can efficiently handle NGS data (Paris et al., 2017). Several pipelines, with differing characteristics, are available for processing data, including those for reduced representation methods. These include; RaPID (Willing et al., 2011), RADtools (Baxter et al., 2011), Rainbow (Chong et al., 2012), Stacks (Catchen et al., 2013b), dDocent (Puritz et al., 2014), PyRAD

(Eaton, 2014), IPyRaD (Eaton and Overcast, in prep), AftrRAD (Sovic et al., 2015), RADassembler (Li et al., 2018) and RADProc (Nadukkalam Ravindran et al., 2019). Differences between these pipelines, among others, include technical characteristics (e.g. installation requirements, support, speed), pipeline execution (e.g. command-line or command-line interface, parameter space), hardware requirements (e.g. ordinary laptop, server), extra-analysis included (population genetics measures) (Jungwirth, 2017).

2.1.3 Stacks software

One of the most widely-known data-processing software packages for RAD data is Stacks (Catchen et al., 2013b). Stacks is designed to work with any Genotype By Sequencing (GBS) restriction enzyme-based data and provides extensive computational flexibility and documentation, including a full parameter optimisation protocol for the *de novo* assembly of loci (Paris et al. 2017, Catchen et al. 2013a). This protocol involves running a series of *de novo* analyses using different parameter values, evaluating the number of new polymorphic loci present in at least 80% of the samples (r80 rule) until a stable set of main parameter values is found (Paris et al., 2017). The r80 rule is considered to result in the selection of a stable set of loci that are highly replicated across populations, minimising the probability of including paralogous or repetitive sequence, or sequencing errors (Paris et al., 2017). It is also essential during the optimisation process to evaluate fluctuations in the number of loci and SNPs that are scored, which are outcomes of the optimisation procedure (Paris et al., 2017).

The main parameters used to control *de novo* assembly in Stacks are:

- Minimum stack depth (m), which controls the minimum number of raw reads required to form a putative allele.
- Distance allowed between stacks (M), which controls the number of mismatches allowed between stacks to merge them into a putative locus (mismatches are allowed between the two alleles of a heterozygote sample).
- Distance allowed between catalog loci (n), which uses the same algorithm as the M parameter but controls the number of mismatches allowed between any

two alleles of the population (Rochette and Catchen 2017; Catchen et al. 2013b; Catchen et al. 2011).

The optimisation of the main parameter values plays an important role in the data analysis (Mastretta-Yanes et al., 2015), mainly due to biological aspects of the data (level of polymorphism, paralogs, repetitive regions) or by inherent bias to RAD-seq (PCR duplicates, genotype errors). For instance, if the m parameter value is set too low, reads with convergent sequencing error may be erroneously merged. In contrast, if the m parameter value is set too high it is possible that true alleles will not be recorded (Mastretta-Yanes et al., 2015). Thus, correct RAD-seq assembly, using different settings may minimise potential locus-assembly biases and can reduce the problem of under or over-merging reads (Catchen et al., 2013b).

The parameter space that should be explored (for m, M and n) can fluctuate according to the data set. However, a good range of M and n to assess is 1-7, either analysing M and n independently or fixing M=n (Paris et al. 2017; Rochette and Catchen 2017), as both parameters use the same algorithm but at different levels (within and between samples). One recommendation for the m parameter is to set a default value of m=3 (Rochette and Catchen, 2017). However, exploring iterations of m with a range of values from 2 to 5 or even more (depending on the data) has been recommended to gain a broader overview of the behaviour of the data (Paris et al., 2017). Depending on which of the three main parameters are being assessed, the other two parameters can be kept to the default values provided by the Stacks pipeline (Catchen et al. 2013b; see Table 2.1).

Table 2.1: Default main parameter values and core pipeline components of *de novo* assembly of loci using Stacks v2.4

Parameter description	denovo_map.pl	Pipeline component	Default value
Minimum stack depth	m	Ustacks	3
Distance allowed between stacks	M	Ustacks	2
Distance allowed between catalog loci	n	Cstacks	1

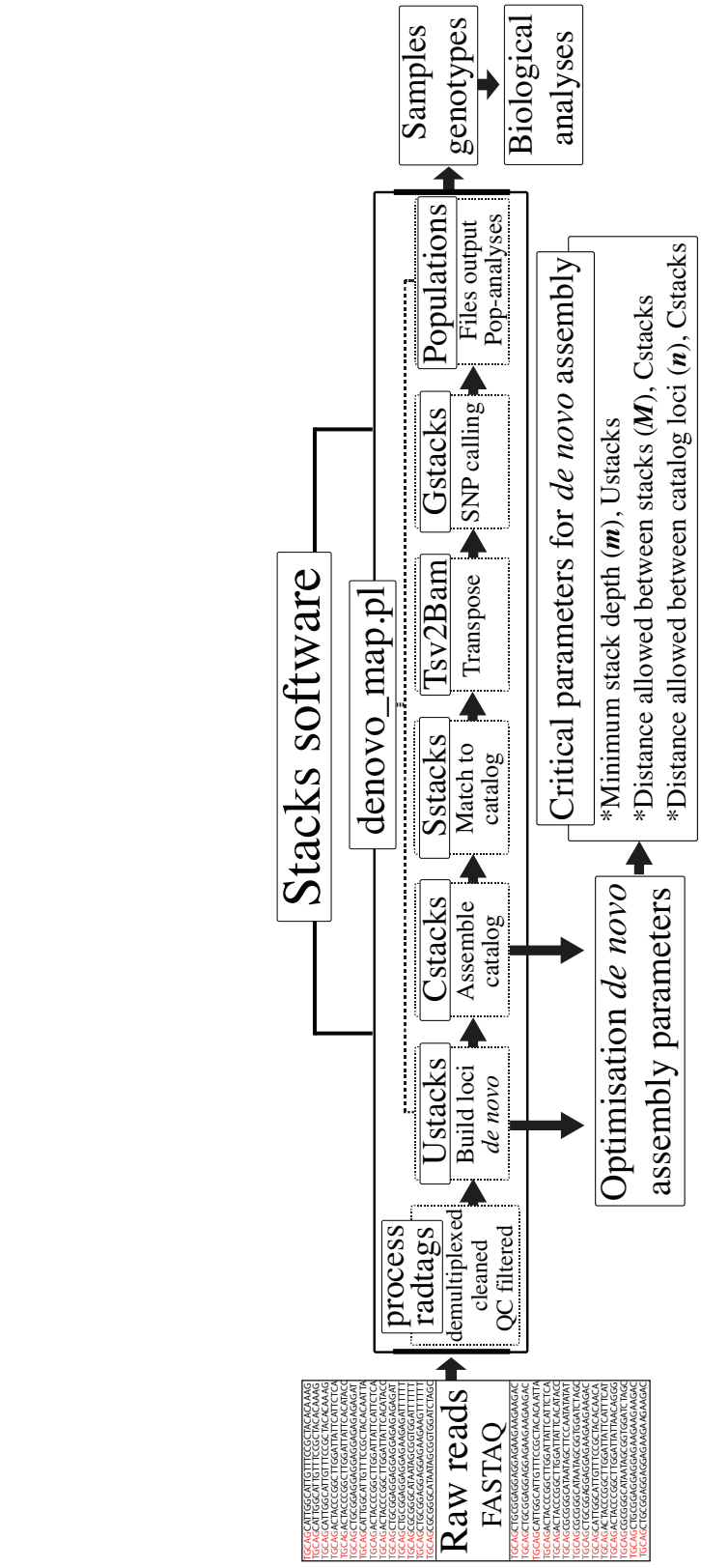


Figure 2.1: Workflow for Stacks software v2.4, showing steps for *de novo* assembly. Samples are first demultiplexed, cleaned, and filtered by quality scores through the *process_radtags* pipeline. Reads are then ready for the *de novo* assembly using the *denovo_map.pl* pipeline which runs as followed; *Ustacks-Cstacks-Sstacks-Tsv2Bam-Gstacks-populations* (*Tsv2Bam* and *Gstacks* are used only with data with forward reads). Each step can be run separately if it needed. Critical parameters for the *de novo* assembly (minimum stacks depth, *m*; distance allowed between stacks, *M*; and distance allowed between catalog loci, *n*) are optimised during the initial steps (in textitUstacks and *Cstacks*)

.

Although the Stacks parameter optimisation protocol is very well defined, only a few studies have rigorously tested it so far and systematically explored the parameter space for a range of settings. Mastretta-Yanes et al. (2015) explored the effect of using different *de novo* assembly settings and evaluated the error rate and the number of polymorphic loci recovered in the non-model plant species *Berberis alpina* Zamudio. They also examined the effect of using different parameter settings on the detection of genetic structure. They suggested that parameter value settings affect the level of error rate and the number of polymorphic loci retained. They also inferred that setting some of the main parameters too high will cause incorrect inference of individual differentiation. A second study, Paris et al. (2017), focused on validating the optimisation parameter protocol for *de novo* assembly of loci, incorporating the r80 rule as a generally useful method to select the core parameters for Stacks. This study utilised three different organisms; brown trout (*Salmo trutta* Linnaeus), king penguin (*Aptenodytes patagonicus* Miller) and red earthworm (*Lumbricus rubellus* Hoffmeister). Finally, Díaz-Arce and Rodríguez-Ezpeleta (2019) optimised *de novo* assembly parameters to evaluate the effects of read-filtering, locus assembly and polymorphic site selection on the number of markers obtained and genetic differentiation inferred. This study employed three different species; green crab (*Carcinus maenas* Linnaeus), Atlantic mackerel (*Scomber scombrus* Linnaeus) and Atlantic deep-sea scallop (*Placopecten magellanicus* Gmelin). They concluded that recovering a higher number of polymorphic loci is not necessarily associated with higher genetic differentiation. They also concluded that the presence of PCR duplicates, selected loci assembly parameters and selected SNP filtering parameters affect the number of recovered polymorphic loci and degree of genetic differentiation.

2.1.4 Objectives

Given the importance of the pipeline and the parameter settings that control *de novo* assembly of loci in non-model species, in the current study, I focus on *de novo* locus assembly, paying particular attention to enhancing our understanding of the impact of parameter value optimisation in conifer species for the first time. Specifically, I assess the optimisation protocol recommended by Paris et al. (2017) and Rochette and Catchen (2017) in four non-model plant species with large genomes and evaluate the sensitivity of parameter settings on locus assembly and population genetic statistics.

2.2 Material and Methods

2.2.1 Sample collection

Leaf tissues (needles) were collected during the austral summer (2017) from a total of 113 individuals of South American endemic conifer species: three Podocarpaceae species; *Saxegothaea conspicua* Lindley (29 individuals, from 9 populations), *Prumnopitys andina* (Poepp. ex Endl) de Laubenfels (30 individuals, from 8 populations), *Podocarpus salignus* D. Don (30 individuals, from 8 populations) and one member of the Cupressaceae family; *Fitzroya cupressoides* (Molina) Johnston (24 individuals, from 8 populations).

Samples were collected from the entire natural range of each species in Chile, and individuals were randomly selected with at least 50 m between trees to minimise sampling closely-related individuals. Leaf tissue (needles) was stored in silica gel after collection.

2.2.2 DNA extraction and sequencing methods

DNA was extracted from leaf tissues using the DNeasy^{TM} Plant Minikit, (QIAGEN, Edinburgh, UK), following the manufacture's protocol[1]. Minor modifications to the standard protocol (described below) were made to obtain the DNA concentration required for RAD-seq. The DNA quality was assessed on an agarose gel, and DNA concentration was measured using a Qubit 2.0 (Invitrogen ®). Samples for which we obtained concentrations higher than 20 ng/ µL were diluted with AE Buffer to normalise all samples to the same concentration. Restriction site Associated DNA sequencing (RAD-seq) was conducted by the external company Floragenex, USA, who provided the raw data for further analyses.

Modification of DNeasy^{TM} Plant Minikit standard protocol

(i) **Sample preparation**

- 30-40 mg dry weight was used instead of 20 mg. recommended.
- Before grinding up the plant material, leaf tissue samples were broken into small pieces with scissors and placed into individual tubes containing two

[1] https://www.qiagen.com

grinding beads (3 mm Retsch cone balls). These were then placed in a freezer at -80 ℃ for 24 hrs, to make it easier to obtain a fine homogeneous powder during grinding.

- A TissueLyser II (also referred to as a Mixer Mill) was used to grind the plant material for a total of 8-10 minutes. The adapter was rotated, and the samples checked every 2-3 minutes to ensure a fine powder was obtained without breaking the tubes through long exposure in the TissueLyser.

(ii) **DNA extraction**: A full, 13-step protocol for DNA extraction from plant material is given in Appendix A. Here we describe only the steps that were modified.

- The mixture was incubated for 30-35 minutes in a freezer at -20 ℃ instead of 5 minutes on ice. This step precipitates detergent, proteins, and polysaccharides; a longer cold period potentially improves protein and polysaccharide precipitation.
- During the process of extracting DNA from the Mini spin column, only 65 µL AE Buffer was used (instead of 100 µL) to increase the final DNA concentration. The elution was incubated for 5 minutes at room temperature (15-25 ℃) before adding another 20 µL of AE Buffer and then incubating for another 5 minutes. After each incubation, samples were centrifuged for 1 minute at 8000 rpm. The total yield obtained was ca. 85 µL per sample.
- DNA was stored in the freezer at -20 ℃.

2.2.3 Library preparation

Box 2.1. Library preparation protocol as used by Floragenex, USA (modified from Baird et al. 2008)

Genomic DNA was digested with the restriction endonuclease PstI and processed into RAD libraries prepared using a similar method to Baird et al. (2008). Briefly, 200 µL of genomic DNA was digested for 60 min at 37 °C in a 50 µL reaction with 20 units (U) of PstI (New England Biolabs [NEB]). After digestion, samples were heat-inactivated for 20 min at 80 °C followed by addition of 2.0 µL of 100 nM P1 Adapter(s), a modified Solexa© adapter (Illumina, Inc.). The PstI P1 adapters each contained a unique multiplex sequence index (barcode), which is read during the first ten nucleotides of the Illumina sequence read. The 100 nM P1 adaptors were added to each sample along with 10x T4 DNA Ligase Buffer (Enzymatics, Inc), T4 DNA Ligase (high concentration [HC], Enzymatics, Inc), and 0.8 µL H2O. The solution was then incubated at room temperature (RT) for 15 min. Samples were heat-inactivated for 20 min at 65 °C, pooled, and randomly sheared with a Bioruptor (Diagenode) to an average size of 500 bp. Samples were then run out on a 1.5% agarose (Sigma), 0.5X TBE gel, and DNA fragments of 200 bp to 400 bp were isolated using a MinElute Gel Extraction Kit (Qiagen). One reaction of End Repair / dA-Tailing module (NEB) was used to polish the ends of the DNA. After subsequent purification, 0.1 µL of 0.1 µM P2 adapter, a divergent modified Solexa© adapter (Illumina, Inc.), was ligated to the obtained DNA fragments at RT. Samples were again purified and eluted in 15 µL. The eluate was quantified using a Qubit Fluorometer with the dsDNA HS assay kit (Invitrogen), and 10 ng of this product was used in a PCR amplification with 50 µL 2x Phusion Master Mix (NEB), 5 µL of 10 µM modified Solexa© Amplification primer mix (Illumina, Inc.) and up to 45 µL H2O. The amplified library was run on a 1.5% agarose (Sigma), 0.5X TBE gel, and DNA fragments 300 bp to 500 bp were excised and purified as before. The library was quantified with a Qubit fluorometer and run on an Agilent Bioanalyzer with the High Sensitivity kit to determine size, which was 389 bp. 1x100 bp single-end sequencing was performed on the HighSeq.

2.2.4 Quality checking and raw reads

The *process_radtags* pipeline from Stacks v2.4 was executed to examine the raw reads from each library, check the barcode and the RAD cut-site, and demultiplexing the reads. A raw Phred quality score filter of 10 (default value) was applied to remove low quality reads.

2.2.5 Optimisation of the *de novo* assembly of loci using Stacks

To carry out the optimisation protocol for the *de novo* assembly of loci using the Stacks V2.4 pipeline, we followed the recommendations of Paris et al. (2017) and Rochette and Catchen (2017). This protocol consists in exploring the settings of the three main parameters that control *de novo* assembly (m, M and n), using the r80 rule (including a minimum of 80% of individuals in a population required to process a locus for that population), as discussed in more detail above. Because our samples comprised only 2-5 individuals per population, all samples within each data set were treated a single population.

Optimising minimum stack depth (m)

FASTQ files from each data set were processed using the *denovo_map.pl* pipeline. The m parameter was explored with a range of values from 2 to 5 for each data set, evaluating coverage, number of SNPs, number of assembled loci and number of polymorphic loci obtained for each value of m. The other main parameters that control the *de novo* assembly of loci in Stacks (M and n) were set to default values.

Optimising distance allowed between stacks (M) and distance allowed between catalog loci (n) values

Two approaches were used to select the parameter values M and n. In both methods, the m parameter was fixed to the default value m= 3.

(i) The first approach focused on analysing both parameter values (M and n) independently, exploring each within the range 1-7 for each species, evaluating the

number of SNPs, number of assembled loci and number of polymorphic loci for each iteration of M and n.

(ii) The second approach fixed both parameters (M and n) with the same value, $M{=}n$ (Mn), following the recommendations of Paris et al. (2017) and Rochette and Catchen (2017). Values tested for *Mn* ranged from 1-7, surveying only the number of new polymorphic loci added for each iteration of *Mn*.

2.2.6 Sensitivity of *de novo* locus assembly to optimised parameter values

After defining optimised parameter values for *de novo* locus assembly of each data set, two different approaches were used to evaluate the impact and sensitivity of the process to these parameter settings. In contrast to the previous analyses which treated all samples from a given species as a single population, we explored the data assigning each sample to its corresponding population, incorporating in this analysis only loci that were present at least in 80% of the total number of samples, using the "R" filter (minimum percentage of individuals across populations required to process a locus).

(i) The first approach explored the sensitivity of population genetic analyses to parameter settings, exploring the population genetics measures F_{ST} and the level of nucleotide diversity (pi). The *denovo_map.pl* (including the *population* pipeline) was executed using both default and optimised parameters value for each data set. The global F_{ST} value was determined using the diveRsity package in R (Keenan et al., 2013), using the diffCalc function, based on Weir and Cockerham (1984). The pi value was recorded from each population in each data set using the *populations* pipeline in Stacks. The pi calculation was defined in Stacks using equation 2 from Hohenlohe et al. (2010).

(ii) The second analysis explored the impact of missing data on the number of assembled loci, polymorphic loci and SNPs. We executed the *denovo_map.pl* assembly pipeline using the optimised parameter values found for each data set. Three different levels of missing data were allowed in each data set using R filters: 1) R40, which includes a minimum of 40% of individuals across populations to process a locus, retaining up to 60% of individuals with no data for that locus; 2) R60, which includes a minimum of 60% of individuals across populations to process a locus; and 3) R80, which includes a minimum of 80% of individuals across populations to process a locus. The number of total sites, polymorphic sites and variant sites (SNPs) retained by each iteration of missing data were recorded from the output file of the *de novo* assembly (*denovo_map.pl*). Global F_{ST} values were also assessed using the diveRsity package in R (Keenan et al., 2013), using the diffCalc function.

2.2.7 Plotting data

All analyses were plotted using the ggplot2 package in R V3.4.4 using the interface R studio V1.1447, and Illustrator CC 2018.

2.3 Results and Discussion

2.3.1 Chromosome number and genome size of the conifer species

Conifers usually have chromosome numbers in the range of 2n= (14) 20-24 (38); however, *F. cupressoides* is a tetraploid species with a chromosome number of 2n= 4x= 44. *Saxegothae conspicua* is a diploid species with a chromosome number 2n= 24. *Prumnopitys andina* has a chromosome number of 2n= 38 (determination from a cultivated plant) (Hair and Beuzenberg, 1958). However, it is uncertain if the species is diploid, though, the closer taxa in its phylogeny show similar numbers of chromosomes, therefore is also considered most likely to be a diploid species (Zonneveld, 2012). *Podocarpus salignus* has a chromosome number of 2n= 38, yet it is also uncertain if the species is diploid. The closer taxa in its phylogeny show similar numbers of chromosomes. However, other *Podocarpus* species have fewer chromosomes (ranging 2n= 20-24, is possible it is an ancient tetraploid) (Hair, 1966). Conifer species often have large genomes, with large amounts of repetitive DNA (Pellicer et al., 2018). Conifer genome size expressed as nuclear DNA content (2C-value) can vary from 8.3 to 71.6 picogram (pg) (Zonneveld, 2012), with 1 pg. equivalent to 978 megabases (Dolezel et al., 2003). The genome sizes of Podocarpaceae species are smaller than other conifers, with 2C-values of only 8-28 pg (Zonneveld, 2012). A direct measure is available for *S. conspicua*, which shows a 2C-value of 10.2 pg. 2C-values were not available for *Pr. andina* and *P. salignus*, however, the genera show average 2C-values of 11 and 14-17 pg respectively. The genome size of *F. cupressoides* has been estimated as 2C-value= 35 pg (Zonneveld, 2012).

2.3.2 DNA concentration

Using the DNeasy^{TM}Plant Minikit, a total of 112 from 113 samples yielded DNA concentrations of more than 20 $\pm$ 3 ng/ µL, as required for RAD-seq (65 µL per sample) (See Fig 2.2). Overall, the small modifications made to the standard DNeasy^{TM}Plant Minikit protocol (increasing the amount of plant material used, freezing the leaf tissues before grinding, increasing the length of cold incubation during the precipitation stage and reducing the quantity of AE buffer used in the final elution step), gave good results for the extraction of DNA from the four conifer species, yielding DNA concentrations

significantly more than those required for RAD-seq in almost all samples. Our modification to increase the initial amount of plant material might appear counter-intuitive, as exceeding the recommended amount of starting material can reduce yield and DNA purity. However, we found that initial DNA extractions using 20 mg dry weight plant material were not successful while increasing the amount of needle tissue led to adequate DNA concentrations.

2.3.3 Number of reads and coverage

Number of reads

More than 95% of the total initial raw reads in each species were retained after running the *process_radtags* pipeline. *Saxegothae conspicua* retained the most mean reads per sample after filtering, with a total of 9 million reads, representing 97.6% of the total initial raw reads for samples in this species. This was followed by *Pr. andina* with 7.7 million, equivalent to 98.2%, *P. salignus* with 6 million reads (96.9%), and *F. cupressoides* with 6.8 million reads (96.5% of total initial raw reads; Figure 2.3). Overall, between 2 and 4% of total raw reads in each data set were discarded for low quality. This may suggest a successful sequencing performance for the four conifer species data sets or may be the consequence of applying a relaxed sequencing filter (Q10). The Stacks module *process_radtags* uses a sliding window down the length of the reads and checks the quality score within this window; thus if the score falls below the probability assigned, the read is discarded [2]. In this case (using the Phred quality score of 10), all the reads that scored below 90% probability of being correct were discarded.

[2] http://catchenlab.life.illinois.edu/stacks/

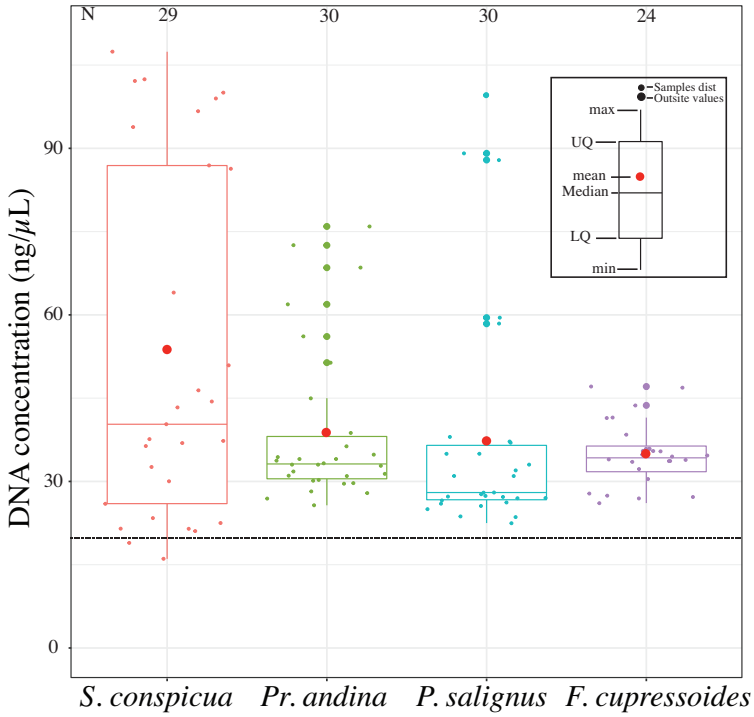


Figure 2.2: Distribution of DNA concentrations obtained for the conifer species (*S. conspicua*, *Pr. andina*, *P. salignus* and *F. cupressoides*) following DNA extraction using the DNeasy^{TM} Plant Minikit. Dashed line indicates the minimum DNA concentration required to carry out the RAD-sequencing. Numbers above each boxplot (N) indicate the number of samples in each data set. Red dots indicate the mean DNA concentration for each species. Full list of DNA concentration by sample and species can be consulted in Appendix A.1.

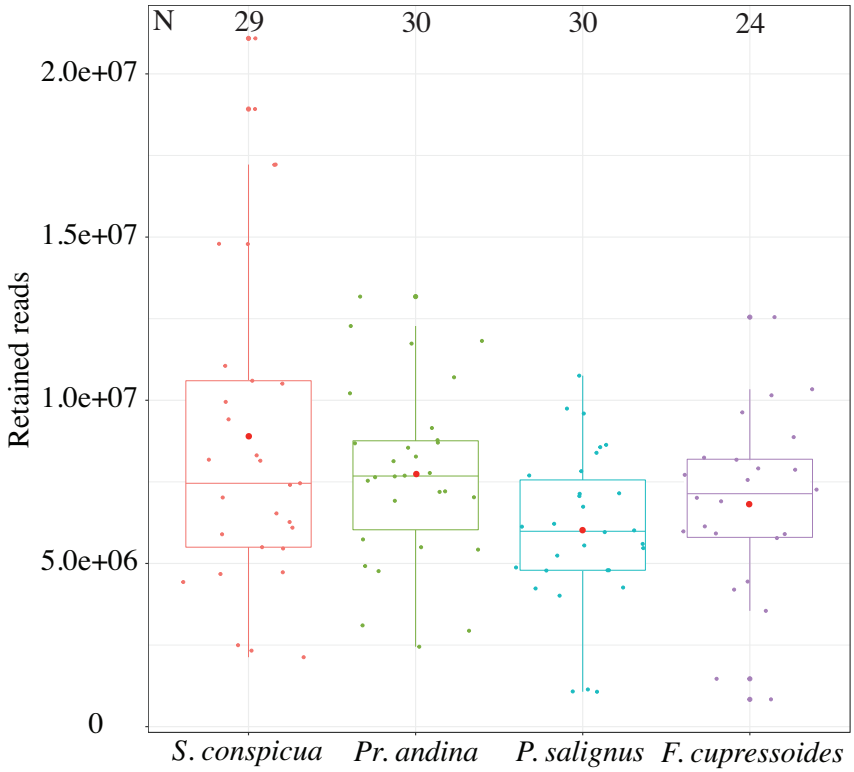


Figure 2.3: Distribution of reads retained after applying *process_radtags* in Stacks on each data set (*S. conspicua*, *Pr. andina*, *P. salignus*, and *F. cupressoides*). Numbers above each boxplot (N) indicate the number of samples in each data set. Red dots indicate the mean DNA concentration for each species.

Coverage

Generally, coverage improved slightly as the value of m increased, for all data sets (Figure 2.4). The most highest mean coverage per sample (merged reads) was found in the m5 iteration with a total of 96.8X ± 20.9 in *Pr. andina* followed by *F. cupressoides* with 91.2X ± 3.2, *P. salignus* with 89.8X ± 23.4, and *S. conspicua* with 86.2X ± 23.6. The lowest coverages for merged reads were found with the m2 setting, with five samples showing coverage ranging from 24 to 28X. Despite this, the final mean coverage for each data set in each iteration of m did not drop below 76X, suggesting a low level of contamination in the four data sets and sufficient coverage to conduct further population genetics analyses (Kofler et al., 2011).

This result is consistent with the Paris et al. (2017) where, in three biologically different data sets (a fish, *S. trutta*; a penguin, *A. patagonicus*; and an earthworm, *L. rubellus*) the mean merged coverage increased slightly with increasing m value. In our case, increasing coverage by increasing the value of m may be explained by the percentage of secondary reads incorporated into the primary stack depth (gap allowing by merging secondary reads (N) or M+2) which was slightly increased by increasing m (for further information about the stack formation see Appendix A.1 and A.2).

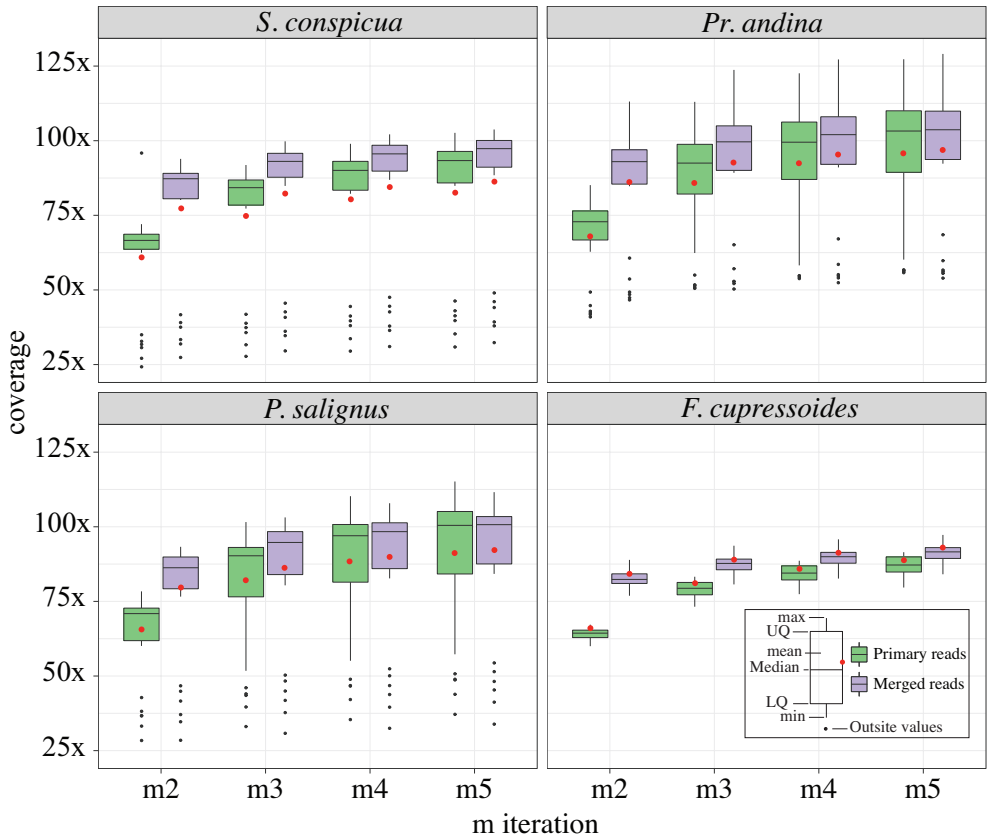


Figure 2.4: Coverage distribution for *S. conspicua*, *Pr. andina*, *P. salignus* and *F. cupressoides* primary and merged reads. Green bars indicate coverage for primary reads and purple the coverage after merging alleles by each iteration of *m*. Coverage was obtained from *denovo_map.pl* assembly, modifying the minimum number of reads (*m*) used to create a stack, and using default values for the other main parameters (*M* and *n*).

2.3.4 Optimising minimum stacks depth (*m*) value

Generally, the number of assembled loci, polymorphic loci and SNPs increased sharply when *m* was increased from m2 to m3, in three of the four data sets (*S. conspicua*, *Pr. andina* and *P. salignus*; Figure 2.5). Above m3, the number of polymorphic loci increased slightly until it stabilised at m5 for *S. conspicua* and *P. salignus* and m4 for *Pr. andina*. In contrast, in the *F. cupressoides* data the number of assembled loci, polymorphic loci and SNPs remained almost constant for each *m* iteration. The general trend found is congruent with Paris et al. (2017) and the more recent findings of Díaz-Arce and Rodríguez-Ezpeleta (2019). Both studies observed that the number of polymorphic loci increased when *m* was increased from 2 to 3 and, after reaching a peak, the number of polymorphic loci remained constant or decreased slightly for

further increases in m value. Díaz-Arce and Rodríguez-Ezpeleta (2019) suggest that the reduction in the number of shared loci after reaching a peak at a specific m value could be attributed to the level of missing loci, making it more difficult for a locus to be shared between individuals. In our study, the most polymorphic loci was found in *Pr. andina*, with 888 loci, followed by *P. salignus* with 532 loci, *S. conspicua* with 386 loci and *F. cupressoides* with considerably fewer loci, 51. This low number of assembled loci and polymorphic loci in *F. cupressoides* (after applying the r80 filter) could be the result of retaining the majority of the loci only in a few samples, possibly the consequence of a high level of incomplete DNA digestion in this data set.

Ultimately, optimised m values were established for *S. conspicua* and *P. salignus* at m5 and for *Pr. andina* and *F. cupressoides* at m4 (Figure 2.5), with these values maximising the number of assembled loci for each species. This differs from the recommendation of Paris et al. (2017) and Rochette and Catchen (2017) to use the default value, m3. Nevertheless, the difference is small: the optimised m values found for our datasets retained only a few more polymorphic loci compared to m3 (77 new loci for *P. salignus*, 44 for *S. conspicua*, 15 for *Pr. andina* and one for *F. cupressoides*).

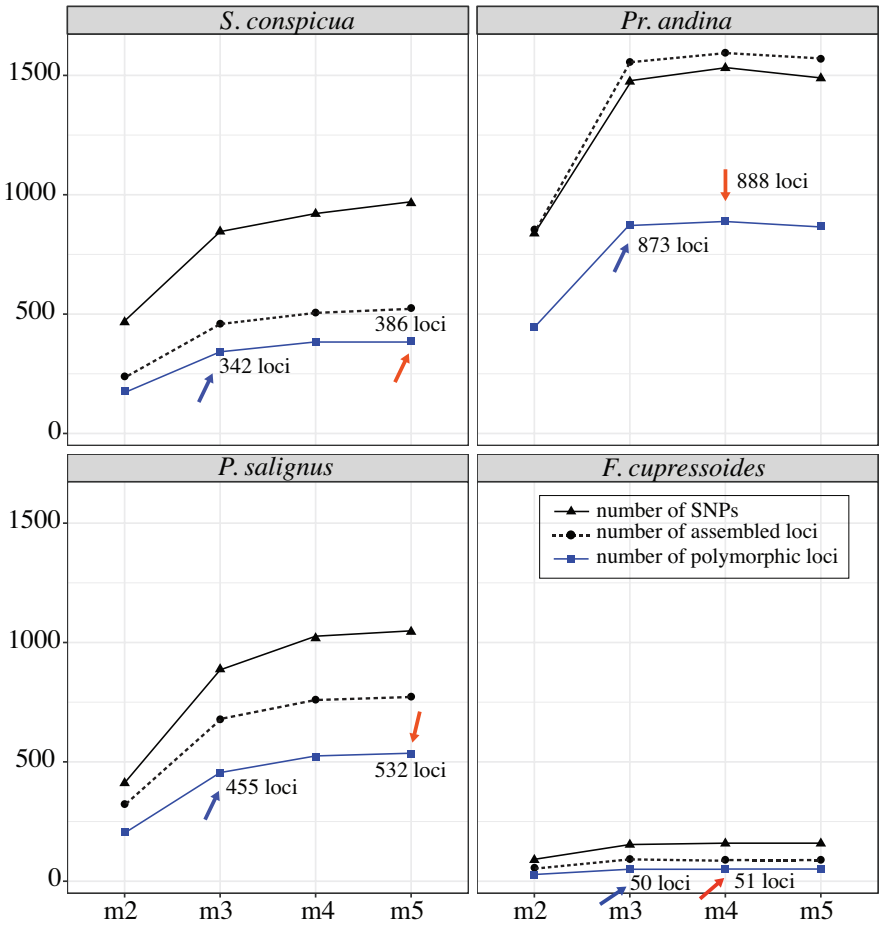


Figure 2.5: Plots illustrating the relationship between the minimum number of reads required to create a stack (m) and; a) number of SNPs b) number of assembled loci and c) number of polymorphic loci. Red arrows indicate the maximum number of polymorphic loci in each data set, representing the optimised m value (*S. conspicua*=m5, *Pr. andina*=m4, *P. salignus*=m5, *F. cupressoides*=m4). Blue arrows indicate the number of polymorphic loci added using the m3 value recommended by Paris et al. (2017)

.

2.3.5 Optimising values for the distance allowed between stacks (M) and the distance allowed between catalog loci (n), handling both parameter values independently

Generally, the number of assembled loci, polymorphic loci and SNPs increased as the M value increased (Figure 2.6). Only the *F. cupressoides* data set showed little impact of increasing from M1 to M7, with a slight fluctuation between 47-58 polymorphic loci. Our results partly contrast with those of Paris et al. (2017), where the general pattern found was that higher values of M led to decreasing numbers of assembled loci, but show

a similar pattern in terms of number of polymorphic loci and SNPs, which increased as M increased, with no apparent limit in the average amount of polymorphism detected. This may be explained by the fact that increasing the M value will merge a higher level of polymorphic alleles within samples (M controls mismatches allowed between the two alleles of a heterozygote sample).

Two of the data sets (*S. conspicua* and *P. salignus*) began to asymptote for the number of polymorphic loci at M4, with a total of 420 and 499 loci respectively. The asymptote for the *Pr. andina* data set began at M6, with 1209 polymorphic loci (Figure 2.6, red arrows). However, these three data sets do not level off completely: the number of polymorphic loci still increases slightly with subsequent increases in M. The largest number of polymorphic loci and SNPs was found in *Pr. andina*, followed by *P. salignus* with 619 loci and 1598 SNPs, and *S. conspicua* with 574 loci and 1741 SNPs, and finally by *F. cupressoides* with 58 loci and 211 SNPs at M7. This result suggests that each data set contains a small number of loci with a high-density of SNPs, similar to that of Paris et al. (2017).

Similarly, the general pattern for n was that increasing n values corresponded to increases in the number of assembled loci, polymorphic loci and SNPs. This is mainly because the n parameter value controls the gaps (differences) allowed between samples; thus, increasing the n value will merge a higher number of polymorphic loci across samples. The n parameter plays an essential role in identifying potential polymorphisms: setting the n value too low will restrict the data set to homologous loci in different samples that contain fixed differences. In contrast, setting the n value too high may result in over-merging loci across samples (Paris et al., 2017).

The largest number of polymorphic loci and SNPs occurred at n7 for all data sets, with a total of 1531 loci and 4634 SNPs in *Pr. andina*, 1134 loci and 3812 SNPs in *S. conspicua*, 708 loci and 2066 SNPs in *P. salignus*, and 193 loci and 1075 SNPs in *F. cupressoides* (Figure 2.7). Only two data sets (*F. cupressoides* and *P. andina*) showed signs of stabilisation in the number of polymorphic loci at n4 (i.e., a potential optimised value). In all data sets, additional polymorphic loci were recovered above n4.

In sum, it was hard to find optimised M and n values for these data sets when analysed independently. There were some signs of stabilisation of the number of polymorphic loci using M2 and n4 for *F. cupressoides* and M4 and n4 for *P. salignus*; however, for every higher setting of each parameter we still found new polymorphic loci. The other two data sets (*S. conspicua* and *Pr. andina*) did not show any clear stabilisation in number of SNPs, number of assembled loci or number of polymorphic loci through increasing M and n parameters independently.

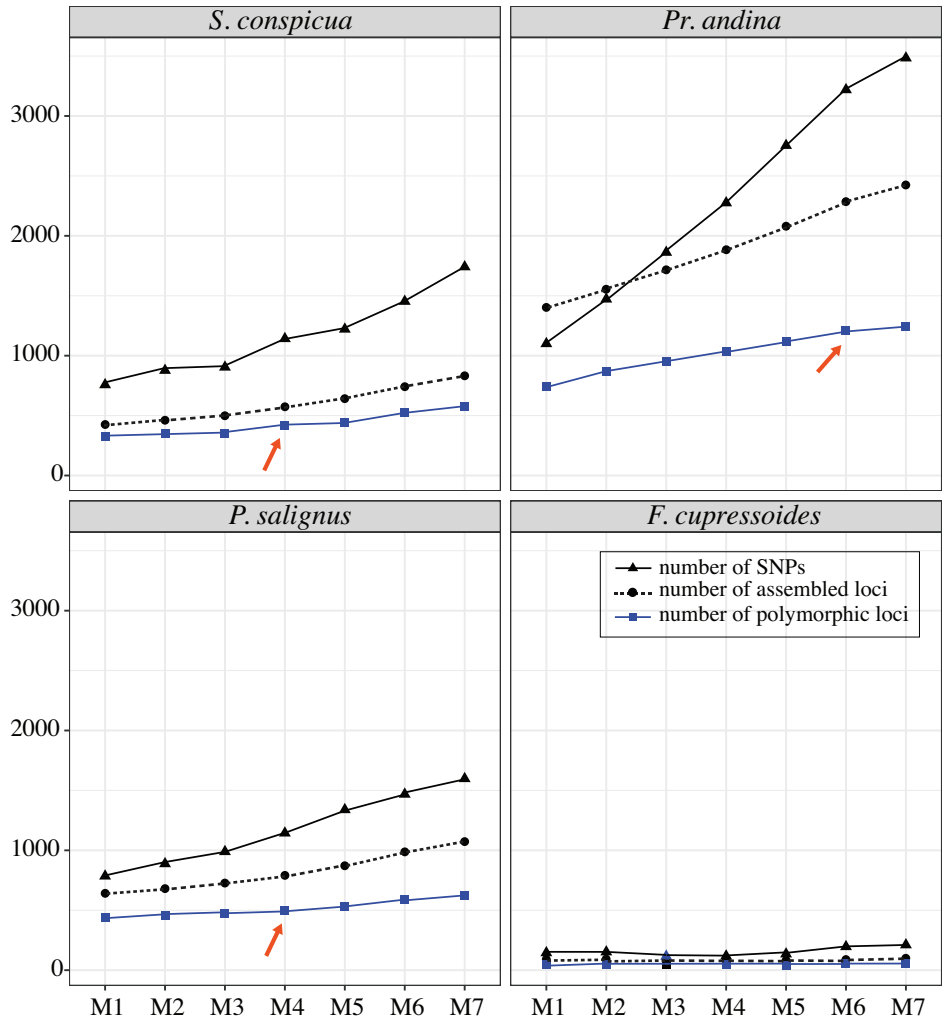


Figure 2.6: Impact of changing the distance allowed between stacks (M) on the numbers of SNPs, assembled loci and polymorphic loci; m and n parameters were kept at default values. The plot shows that increasing M results in increasing the numbers of SNPs, assembled loci and polymorphic loci. Red arrows indicated the potential optimised M values for three of the four data set (M4 for *S. conspicua*, possible M6 or M7 for *Pr. andina*, M4 for *P. salignus*

.

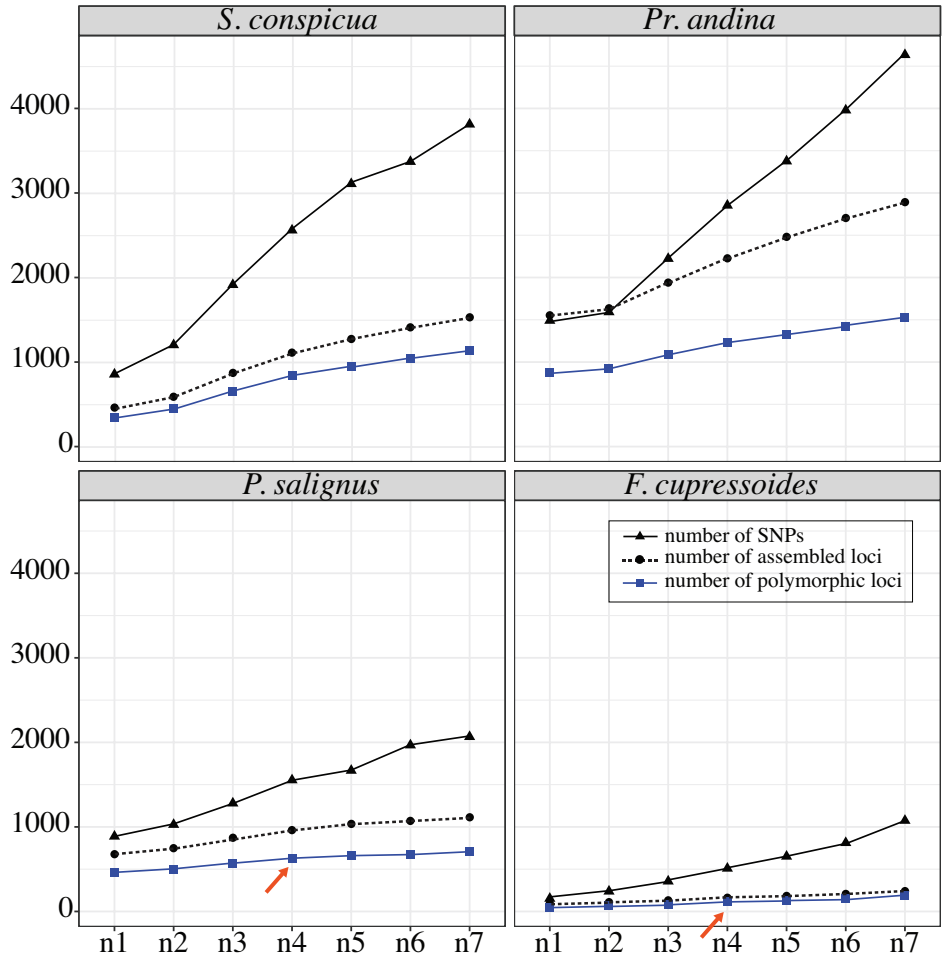


Figure 2.7: Impact of changing the distance between catalog loci (n) on the number of SNPs, assembled loci and polymorphic loci; m and M parameters were kept at the default values. Red arrows indicate the potential optimised values.

2.3.6 Optimising values for the distance allowed between stacks (M) and the distance allowed between catalog loci (n), while fixing the two parameters to the same values

Across all four species, we found little change in the number of new polymorphic loci found with increasing values of *Mn*. Only *Pr. andina* showed the most significant amount of new polymorphic loci at Mn1, with fewer new loci at Mn2 but approximately similar numbers of new loci from Mn2 upwards (Figure 2.8). The other three species (*S. conspicua*, *P. salignus* and *F. cupressoides*) displayed roughly the same number of new polymorphic loci with each *Mn* iteration. However, at Mn5 the number of new polymorphic loci almost doubled compared to Mn4 for two species (*P. salignus* and *F.*

cupressides). Since our results showed no *Mn* iteration that resulted in fewer loci (which would appear as negative values in Figure 2.8), in theory, it would be possible to select different *Mn* values for each data set. However, to minimise any bias, our selection of a parameter value for *Mn* focused on identifying the first "dip" using a stringent criterion in order to maximise merging of homologous loci and reduce retaining over-merged loci across samples, which may provide the balance between obtaining potential true polymorphic loci and introducing sequencing error (Paris et al., 2017).

This first dip can be seen at Mn4 for *S. conspicua* and *P. salignus*, the latter species showing a similar result to that obtained by analysing both parameters (*M*and *n*) independently. At Mn4, *S. conspicua* shows a total of 655 polymorphic loci and 2154 SNPs and *P. salignus* 603 polymorphic loci and 1587 SNPs. For *Pr. andina* we selected an optimised value of Mn3, with a total of 1041 polymorphic loci and 2134 SNPs, and for *F. cupressoides* Mn2 was indicated as the potential optimised value, with only 66 polymorphic loci and 252.

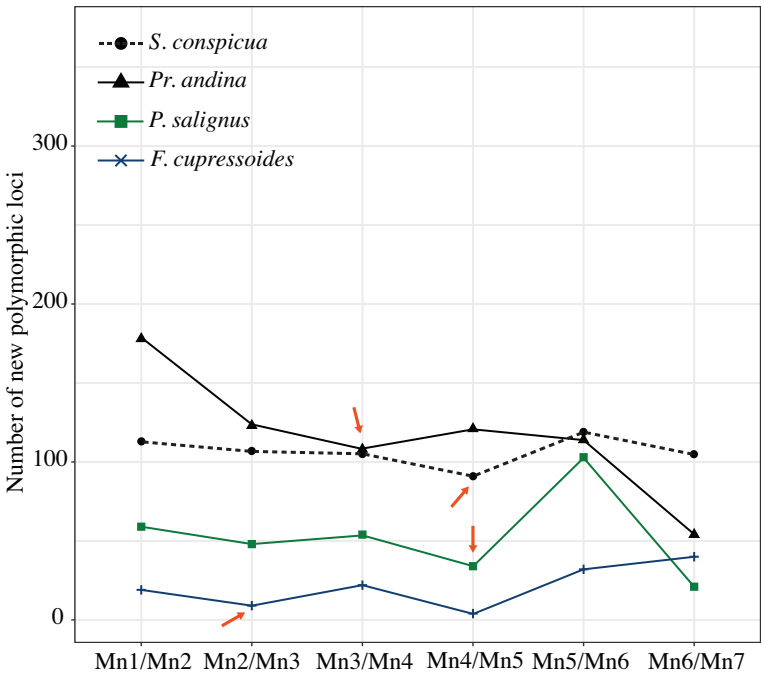


Figure 2.8: Number of new polymorphic loci added for each iteration of *M*=*n* (*Mn*) for the four data sets (*S. conspicua*, *Pr. andina*, *P. salignus* and *F. cupressoides*). Red arrows indicate Mn iteration selected as representing the optimised Mn parameter value in each data set. (*S. conspicua*, Mn=4; *Pr. andina*, Mn=3; *P. salignus*, Mn=4 and *F. cupressoides*, Mn=2)

2.3.7 Optimisation of the *de novo* parameter values for the four data sets

Results of the optimisation protocol, recommended by Paris et al. (2017), revealed that optimised main parameter values (m, M and n) differ between species and also from the default values defined by the Stacks software (see Table 2.2). This result is congruent with previous studies (e.g., Paris et al. 2017, Mastretta-Yanes et al. 2015). These differences between data sets could be attributed to the optimised parameter values depending upon the polymorphism of the genome being analysed, the level of sequencing error and the depth of sequencing performance (Mastretta-Yanes et al., 2015). On the other hand, the general pattern shows that the main parameter values selected for the *de novo* assembly of loci correspond to the largest number of assembled loci, polymorphic loci and SNPs for all data sets (Figure 2.9). We consider that the effect of performing an optimisation protocol may be to provide a sufficient balance of retaining potential true polymorphic loci and SNPs versus sequencing errors (Paris et al., 2017). The most significant differences were found in the number of SNPs identified after applying the optimised values to the *de novo* assembly of loci. *S. conspicua* displayed a total of 2068 SNPs, 1222 more SNPs than using the default values; *Pr. andina* showed a total of 2246 SNPs, 799 more SNPs than using default values; *P. salignus* a total of 1660 SNPs, 772 more SNPs than using default values; and *F. cupressoides* 256 SNPs, 103 more SNPs than using default values. Regarding the number of polymorphic loci, the biggest difference was also found in *S. conspicua*, with a total of 663 loci, 321 more than using the default value. *Podocarpus salignus* had a total of 1660 loci, 207 more loci than with the default values; *Pr. andina* a total of 1068 polymorphic loci, 195 loci more than with default values; and *F. cupressoides* 74 loci, 24 more loci than with default values.

Table 2.2: Optimised main parameter values selected for the *de novo* assembly of loci for each conifer species

Data set (species)	m	M	n
S. conspicua	5	4	4
Pr. andina	4	3	3
P. salignus	5	4	4
F. cupressoides	4	2	2

default parameter values: m=3, M=2 and n=1

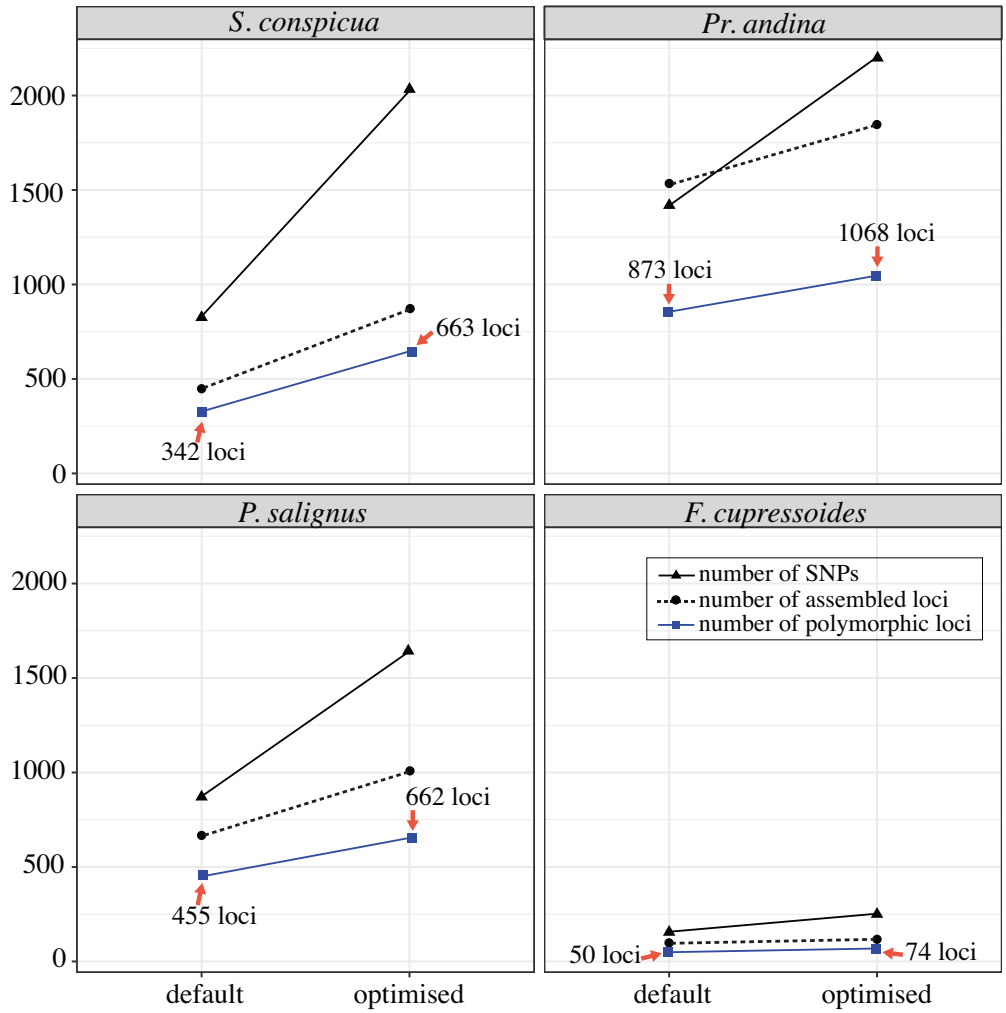


Figure 2.9: Plot showing the differences between the default (m3, M2 and n1) and optimised values of the *de novo* assembly of loci by data set (*S. conspicua*: m5, M4 and n4; *Pr. andina*: m4, M3 and n3; *P. salignus*: m5, M4 and n4 and finally *F. cupressoides*: m4, M2 and n2), for the number of SNPs, number of assembled and polymorphic loci using the r80 rule. Each data set was treated as a single population (*F. cupressoides* with 29 samples, *Pr. andina* with 30 samples, *P. salignus* with 30 samples and *F. cupressoides* with 24 samples)

2.3.8 Sensitivity analysis of using the optimised parameters values of the *de novo* assembly of loci

Sensitivity of population genetics statistics to the optimised parameter settings

Global F_{ST} values remained almost the same across default and optimised parameter values for three of our four data sets (species), increasing slightly only in *S. conspicua*, from 0.037 to 0.049 (see Figure 2.10). Global F_{ST} also decreased slightly in *F. cupressoides* when optimised parameter settings were applied compared to default values; however, for this species, both F_{ST} values (default and optimised) were not statistically significantly different from zero. This result differs from Mastretta-Yanes et al. (2015), who compared mean F_{ST} values (based on the mean pairwise values) in *Berberis alpina*. They observed an increase in mean F_{ST} from 0.07 (using default parameter settings) to 0.19 (using optimised parameter settings).

The levels of nucleotide diversity during *de novo* assembly of loci, measured by pi values, increased slightly in three out of four data sets (*Pr. andina*, *P. salignus* and *F. cupressoides*) using optimised parameter settings compared to the default settings. The most significant difference in mean pi value was found in *Pr. andina*, increasing from 0.0027 to 0.0035, followed by *P. salignus* (0.0022 to 0.0026) and *F. cupressoides* (0.0011 to 0.002). This expected pattern in pi value, increasing diversity by increasing the distance in the parameter spaces (Table 2.2), may be explained by the fact that an increase in the number of mismatches allowed between sample loci (n) during catalog formation in Stacks will increase the average number of nucleotide differences per site. In contrast, the mean pi value for the *S. conspicua* data set decreased slightly from 0.0027 to 0.0025 (Figure 2.11).

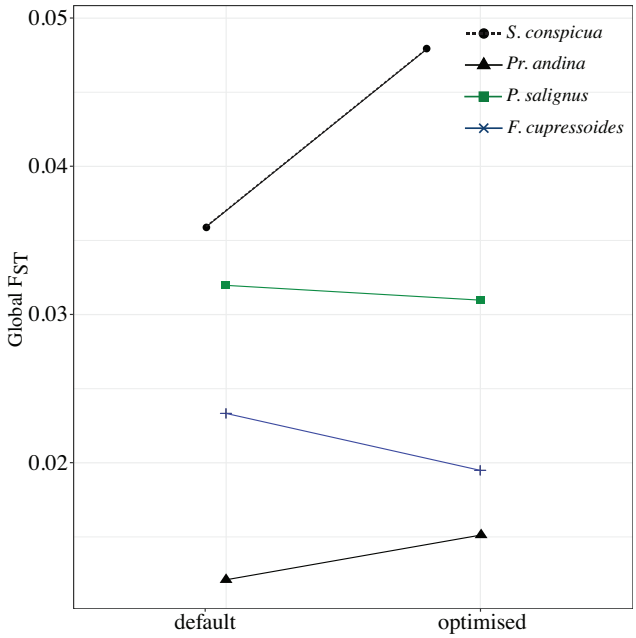


Figure 2.10: Comparison of the effect on F_{ST} values of using default vs. optimised parameters for *de novo* assembly of loci after applying an R filter (80% of individuals across population required to process a locus).

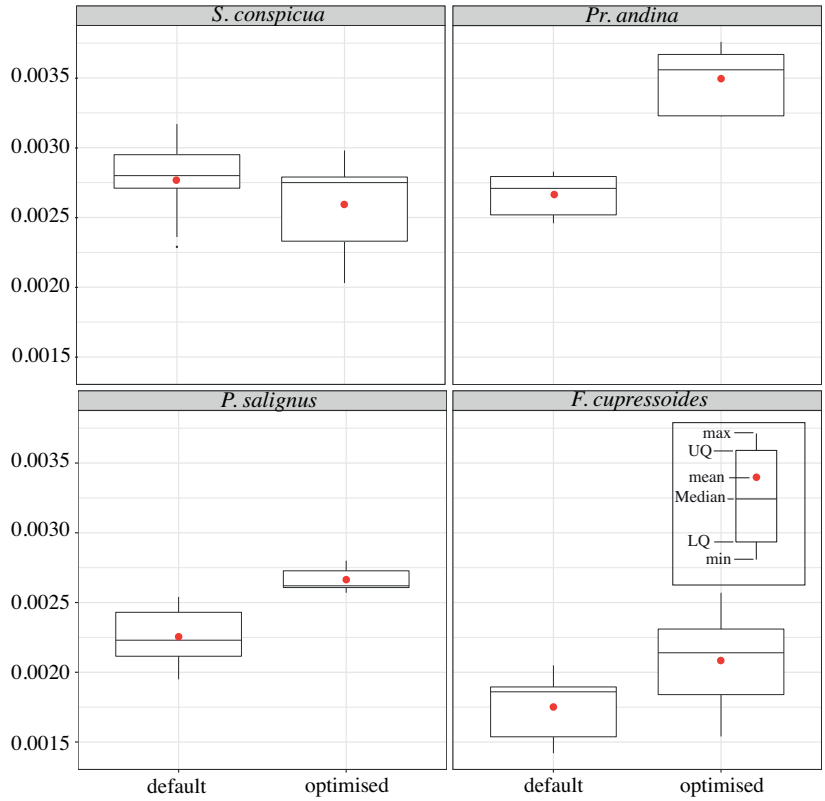


Figure 2.11: Comparison of the effect on pi value of using default vs. optimised parameter values in *de novo* assembly of loci after applying an R filter (80% of individuals across population required to process a locus).

Sensitivity to the level of missing data

The greater the amount of missing data allowed, the greater the number of sites are included in the analysis: the R80 filter includes only loci that are present in at least 80% of individuals across populations, while the R40 filter, which will include any loci present at least in 40% of individuals, allows a greater number of SNPs. Using the relaxed R40 criterion, the *Pr. andina* data set showed a mean of 7,650,275 total sites, *P. salignus* 3,850,842 sites, *S. conspicua* 3,645,135 sites and *F. cupressoides* 634,345 sites (Figure 2.12). The more stringent R80 filter gave the lowest total number of sites, with a total mean of 170,716 sites for *Pr. andina*, *P. salignus* 92,083, *S. conspicua* 81,492 sites and *F. cupressoides* 10,697 sites. Similarly, the largest number of polymorphic sites and SNPs were found when a higher level of missing data was allowed (R40, Figure 2.13).

Consequently, there is a huge difference in the mean number of polymorphic sites between using a relaxed and stringent filter (R40 and R80): from 25,784 to 460 polymorphic loci in *S. conspicua*, 16,939 to 548 in *P. salignus*, 11,087 to 1,212 in *Pr. andina* and 5318 to 49 in *F. cupressoides*. The number of SNPs in each data set also showed a great difference between using the R40 and R80 filters, with the greatest difference seen in *S. conspicua*. These fluctuations in levels of missing data could be explained by data loss occurring at different stages of the RAD-seq protocol, e.g., during library preparation where mutations at cut-sites could generate null alleles (Gautier et al., 2013) or during data processing and assembly where the coverage threshold and the number of mismatches allowed between loci (n) contribute to the amount of missing data (Mastretta-Yanes et al., 2015).

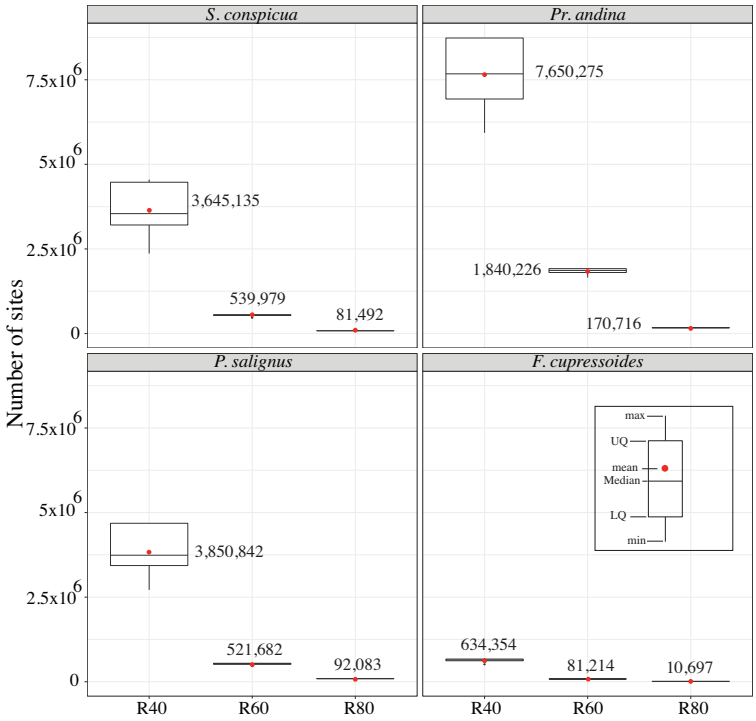


Figure 2.12: Comparison of the total number of sites retained by allowing different levels of missing data (R40, R60 and R80). The number next to each boxplot indicates the mean number of sites obtained across all iterations of R.

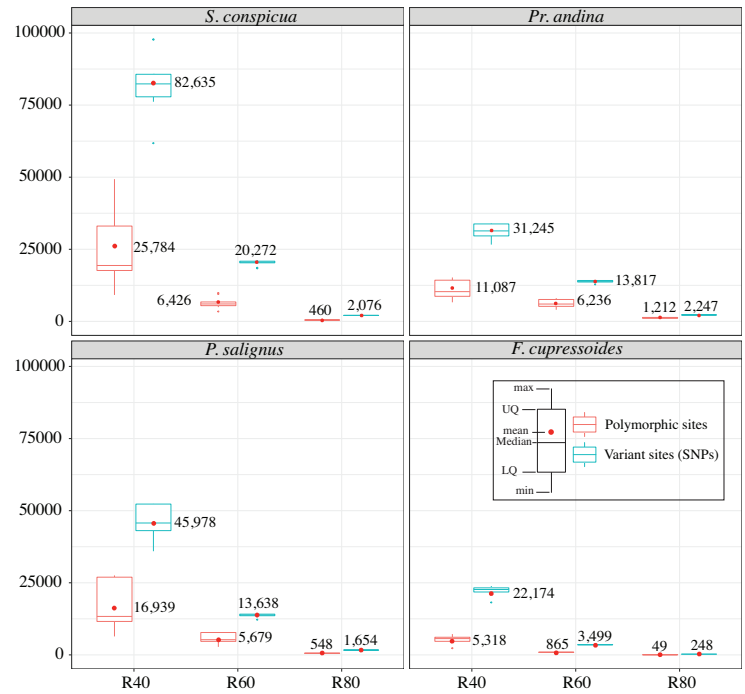


Figure 2.13: Comparison of the total number of polymorphic sites and SNPs retained by allowing different levels of missing data (R40, R60 and R80). The number next to each boxplot indicates the mean number of sites obtained across all iterations of R.

The relationship between global F_{ST} values and the level of missing data did not show a general pattern across the four species (Figure 2.14) and, indeed, the level of genetic differentiation was extremely low and did not differ much when different filters are applied (range: 0.009-0.063). However, we did find that for all four species, the largest global F_{ST} value was found when the greatest proportion of missing data was allowed (R40). This could be the result of incorporating a greater amount of data into these analyses, which might increase bias in the estimation of allele frequencies. In *S. conspicua*, global F_{ST} increased slightly from 0.043 to 0.049 between R60 and R80, while in *Pr. andina*, global F_{ST} decreased with increasing stringency (Figure 2.14). *P. salignus* did not show any variation in global F_{ST} between R60 and R80 (F_{ST}= 0.031). *F. cupressoides* showed little variation in F_{ST} with levels of missing data, with F_{ST} taking almost similar values using R40 and R80 (0.023-0.020 respectively). There is no obvious explanation for the differing patterns in F_{ST} among species, but it seems that calculation of the degree of genetic differentiation is little affected by the levels of missing data which inevitably exist in RAD-seq data sets.

One of the main reason for incorporating a sensitivity analysis of missing data in this investigation is that many of the programs available to evaluate population genetics provide little explanation of how they deal with missing data in genetic statistics calculations. Indeed, in some cases, programs by default fill any empty cells with the average allele frequency observed in the data set (e.g. Genodive). In this scenario, levels of missing data might bias any estimates of genetic structure. In our case, we evaluated the global F_{ST} using the diveRsity package in R (Keenan et al., 2013), using the diffCalc function and based on Weir and Weir and Cockerham (1984). After reviewing several other software packages (e.g. Genodive, Genpop), the R package showed not be influenced by the level of missing data, as it only includes loci that are present in all samples. This lead to us being confident about the further population genetic analyses.

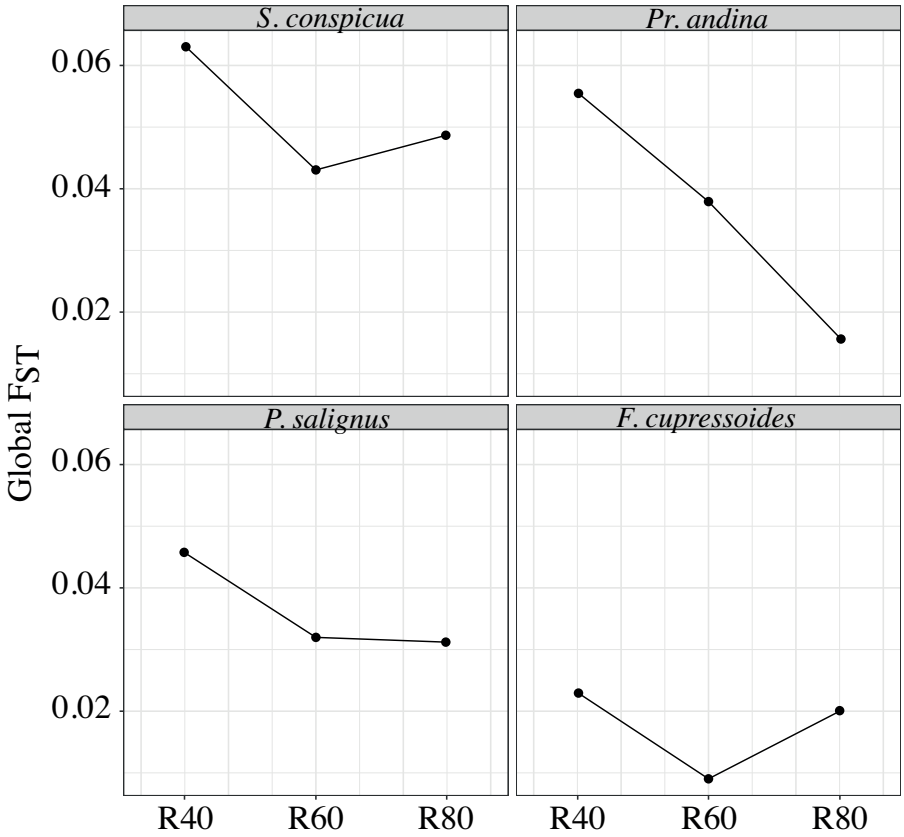


Figure 2.14: Comparison of global F_{ST} values with differing levels of missing data (achieved by R40, R60 and R80 filters) for four conifer species.

2.3.9 Conclusion

In this chapter, we performed *de novo* assembly of RAD loci for four conifer species, exploring the sensitivity of both assembly and population genetic statistics to changing parameter settings. The optimisation protocol provides a framework for exploring and visualising trends in the data used for *de novo* assembly. The full protocol provides information to enable the user to reduce biases during analysis and accurately identify optimised main parameter values for *de novo* assembly. We found that optimised parameter values differed among species, and differed from the default values defined by Stacks. This may be due to the intrinsic biological features of our data sets (e.g., levels of polymorphism) and/or characteristics of library preparation or sequencing performance (e.g., sequencing error, PCR duplicates). Of the three main parameter values that control *de novo* assembly in Stacks, variation in m value had the lowest impact: coverage, number of polymorphic loci and number of SNPs remained almost stable as m was increased above the m3 value recommended by Paris et al. (2017). The selection of values for M and n was clearest when assessing both parameters using the same value, although each Mn iteration independently retained almost the same number of new polymorphic loci, a trend that is repeated in each of the four data sets. Consequently, this analysis reveals that the optimisation protocol recommended by Paris et al. (2017) and Rochette and Catchen (2017) is a very effective way of maximising retention of polymorphic loci and SNPs in these conifer species.

Optimising the *de novo* assembly of loci also affected the inference of population genetic summary statistics: increasing the values of m, M and n to optimised, rather than default values impacted on the population genetic measures explored (F_{ST} and pi). However, the impact in most of the cases was small and is not considered biologically significant. In contrast, the level of missing data allowed in each data set had a great impact on the number of polymorphic loci and SNPs retained. In all analyses, a high level of missing data increased the number of polymorphic loci and, SNPs retained but did not have a significant impact on F_{ST}. It is important to bear in mind that, depending on the software used to calculate at least F_{ST}, the level of missing data could have an impact on calculations of genetic structure.

Based on our evaluation, the following optimised set of parameters (especially for M and n but also for m) have been selected for further study (see next chapter 3).

Optimised parameter values for each species studied

- *S. conspicua*; m5 M4 n4
- *Pr. andina*; m4 M3 n3
- *P. salignus*; m5 M4 n4
- *F. cupressoides*; m4 M2 n2

Bibliography

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*. 3:1–7.

Baxter SW, Davey JW, Johnston JS, Shelton AM, Heckel DG, Jiggins CD, Blaxter ML. 2011. Linkage mapping and comparative genomics using next-generation RAD sequencing of a non-model organism. *PLoS ONE*. 6.

Behjati S, Tarpey PS. 2013. What is next generation sequencing? *Archives of Disease in Childhood: Education and Practice Edition*. 98:236–238.

Birol I, Raymond A, Jackman SD, et al. (24 co-authors). 2013. Assembling the 20 Gb white spruce (*Picea glauca*) genome from whole-genome shotgun sequencing data. *Bioinformatics*. 29:1492–1497.

Borrell JS, Wang N, Nichols RA, Buggs RJ. 2018. Genetic diversity maintained among fragmented populations of a tree undergoing range contraction. *Heredity*. 121:304–318.

Catchen J, Bassham S, Wilson T, Currey M, O'Brien C, Yeates Q, Cresko WA. 2013a. The population structure and recent colonization history of Oregon threespine stickleback determined using restriction-site associated DNA-sequencing. *Molecular Ecology*. 22:2864–2883.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013b. Stacks: an analysis tool set for population genomics. *Molecular Ecology*. 22:3124–3140.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. 2011. Stacks: Building and genotyping loci *de novo* from short-read sequences. *G3: Genes, Genomes, Genetics*. 1:171–182.

Chong Z, Ruan J, Wu CI. 2012. Rainbow: An integrated tool for efficient clustering and assembling RAD-seq reads. *Bioinformatics*. 28:2732–2737.

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*. 12:499–510.

Díaz-Arce N, Rodríguez-Ezpeleta N. 2019. Selecting RAD-Seq data analysis parameters for population genetics: The more the better? *Frontiers in Genetics*. 10:1–10.

Dolezel J, Bartos J, Voglmayr H, Greilhuber J. 2003. Nuclear DNA content and genome size of trout and human. *Cytometry*. 51A:127–128.

Eaton DA. 2014. PyRAD: assembly of *de novo* RADseq loci for phylogenetic analyses. *Bioinformatics*. 30:1844–1849.

Emerson KJ, Merz CR, Catchen JM, Hohenlohe PA, Cresko WA, Bradshaw WE, Holzapfel CM. 2010. Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences of the United States of America*. 107:16196–16200.

Gautier M, Gharbi K, Cezard T, Foucaud J, Kerdelhué C, Pudlo P, Cornuet JM, Estoup A. 2013. The effect of RAD allele dropout on the estimation of genetic variation within and between populations. *Molecular Ecology*. 22:3165–3178.

Hair JB. 1966. Biosystematics of the New Zealand flora, 1945–1964. *New Zealand Journal of Botany*. 4:559–595.

Hair JB, Beuzenberg EJ. 1958. Chromosomal evolution in the Podocarpaceae. *Nature*. 181:1584.

Henson J, Tischler G, Ning Z. 2014. Europe PMC funders group next-generation sequencing and large genome assemblies. *Pharmacogenomics*. 13:901–915.

Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*. 6.

Jungwirth E. 2017. Comparison of ddRAD Analysis Pipelines. Thesis, Graz University of Technology, Graz, Austria.

Keenan K, Mcginnity P, Cross TF, Crozier WW, Prodöhl PA. 2013. DiveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution*. 4:782–788.

Kofler R, Orozco-terWengel P, de Maio N, Pandey RV, Nolte V, Futschik A, Kosiol C, Schlötterer C. 2011. Popoolation: A toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS ONE*. 6.

Li YL, Xue DX, Zhang BD, Liu JX. 2018. An optimized approach for local de novo assembly of overlapping paired-end RAD reads from multiple individuals. *Royal Society Open Science*. 5.

Mastretta-Yanes A, Arrigo N, Alvarez N, Jorgensen TH, Pinero D, Emerson BC. 2015. Restriction site-associated DNA sequencing, genotyping error estimation and *de novo* assembly optimization for population genetic inference. *Molecular Ecology Resources*. 15:28–41.

Nadukkalam Ravindran P, Bentzen P, Bradbury IR, Beiko RG. 2019. RADProc: A computationally efficient *de novo* locus assembler for population studies using RADseq data. *Molecular Ecology Resources*. 19:272–282.

Neale DB, Wegrzyn JL, Stevens KA, et al. (37 co-authors). 2014. Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome Biology*. 15:1–13.

Nowoshilow S, Schloissnig S, Fei JF, et al. (19 co-authors). 2018. The axolotl genome and the evolution of key tissue formation regulators. *Nature*. 554:50–55.

Nunziata SO, Weisrock DW. 2018. Estimation of contemporary effective population size and population declines using RAD sequence data. *Heredity*. 120:196–207.

Paris JR, Stevens JR, Catchen JM, Johnston S. 2017. Lost in parameter space: a road map for stacks. *Methods in Ecology and Evolution*. 8:1360–1373.

Pellicer J, Hidalgo O, Dodsworth S, Leitch IJ. 2018. Genome size diversity and its impact on the evolution of land plants. *Genes*. 9.

Puritz JB, Hollenbeck CM, Gold JR. 2014. dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ*. 2:e431.

Rochette NC, Catchen JM. 2017. Deriving genotypes from RAD-seq short-read data using Stacks. *Nature Protocols*. 12:2640–2659.

Sovic MG, Fries AC, Gibbs HL. 2015. AftrRAD: A pipeline for accurate and efficient *de novo* assembly of RAD-seq data. *Molecular Ecology Resources*. 15:1163–1171.

Vendrami DL, De Noia M, Telesca L, Handal W, Charrier G, Boudry P, Eberhart-Phillips L, Hoffman JI. 2019. RAD sequencing sheds new light on the genetic structure and local adaptation of European scallops and resolves their demographic histories. *Scientific Reports*. 9:1–13.

Weir BS, Cockerham CC. 1984. Estimating F-Statistics for the analysis of population structure. *Evolution*. 38:1358.

Willing EM, Hoffmann M, Klein JD, Weigel D, Dreyer C. 2011. Paired-end RAD-seq for *de novo* assembly and marker design without available reference. *Bioinformatics*. 27:2187–2193.

Zonneveld BJ. 2012. Conifer genome sizes of 172 species, covering 64 of 67 genera, range from 8 to 72 picogram. *Nordic Journal of Botany*. 30:490–502.

Chapter 3

Genetic structure in four endemic South American conifer species with a restricted and fragmented distribution in Chile

3.1 Introduction

3.1.1 Genetic diversity and population structure

Genetic connectivity between populations is essential for the spread of favourable alleles and to replenish genetic variation lost by drift (O'Connell et al., 2007). Where connectivity is disrupted, due to different events or forces (e.g. isolation and/or fragmentation), this can lead to high population differentiation and loss of genetic diversity. The situation where genetic isolation is most likely to occur, reflect a combination of species traits, the landscape in which species occur, and the long term species history (Johnston et al., 2019). Thus higher population differentiation is typically found in species with poor dispersal traits (insect-pollinated, gravity dispersed seeds), short life cycles (and hence susceptible to genetic drift), and/or growing in highly heterogeneous landscapes with lots of barriers to gene flow (Frankham et al., 2004). Significant population differentiation and divergence can also occur where species have been restricted to multiple isolated refugia during past climatic fluctuations, and the signature of long term isolation in refugial populations can persist when expansion occurs from refugia (Marchelli et al., 2010). Over short time scales, population loss (due to habitat destruction and degradation) can also lead to loss of diversity within populations and differentiation between them. This is a significant emerging concern, giving the widescale impacts of human activities on the natural environment, and the importance of genetic diversity for allowing species to respond to environmental change (Nahuelhual et al., 2012).

3.1.2 Population genetic structure in conifers

Usually, conifers are one of the species groups which show low levels of population genetic differentiation (O'Connell et al. 2007, Awad et al. 2014, Hamrick et al. 1992). This is mostly associated with effective pollination mechanism (wind-pollinated), the longevity of individual trees, and species frequently having large population sizes and continuous distributions (at least for conifers from the northern hemisphere). However, there are exceptions to this generalisation. Indeed, several studies have assessed genetic diversity and structure in conifer species with special attention on the effect of isolation and fragmentation. For instance, Tóth et al. (2019) reviewed the impact of isolation

and climatic variability on population differentiation in the long-lived subalpine conifer *Pinus cembra* L., and Eliades et al. (2011) estimated genetic diversity and differentiation in fragmented populations of *Cedrus brevifolia* (Roxb.) G.Don. A common observation from these studies is that range fragmentation and/or spatial isolation have contributed to significant genetic differentiation between populations.

3.1.3 Population genetic structure in the conifers of Chile

The conifer flora of Chile comprises only nine species, but these include 3/8 extant conifer families (Araucariaceae, Cupressaceae, Podocarpaceae) and eight genera, with all of these species and four genera, just restricted to southern South America. The topographical complexity of Chile, and the biology of conifers, in general, lead to different predictions about population genetic structure in Chilean conifers. Thus one prediction leads to a differentiation between populations at a regional scale. The landscape of Chile shows considerable topographical heterogeneity with significant barriers to gene flow created by the mountains of the Andes, the Coastal cordillera, and the intervening Central valley. In addition to the natural barrier to gene flow, human impacts on the environment of Chile have resulted in increased habitat fragmentation. These changes include clearance of land for agriculture and exotic plantations, changes to river drainage and water availability due to hydro-electric dams, and residential and commercial development.

In contrast, the biology of the species is typically associated with high diversity and low differentiation (all are wind-pollinated, and some are incredibly long-lived). However, most of these conifers differ in the way of seed-dispersal (gravity, bird dispersal and wind dispersal, more details in Methods 3.2.2).

3.1.4 Previous studies on population structure in Chilean conifers

To date, only a few population genetics studies have been undertaken on Chilean conifers. e.g. *Araucaria araucana* (Molina) K. Koch (Marchelli et al. 2010, Ruiz et al. 2007, Bekessy et al. 2002), *Austrocedrus chilensis* (D.Don) Pic.Serm. and Bizzarri (Colabella et al. 2014, Pastorino and Gallo 2002), *Fitzroya cupressoides* (Molina) Johnston (Premoli et al. 2000a, Premoli et al. 2000b, Allnutt et al. 1999, Veblen et al. 1976), and *Pilgerodendron uviferum* (D.Don) Florin (Allnutt et al., 2003). These studies detected a range of insights into population genetic structure. For example, some of the authors recognised patterns of genetic differentiation at a regional scale (Andes mountains, Central valley and Coastal range); e.g. Allnutt et al. (1999), Premoli et al. (2000b). Other authors have also observed that genetic distance tended to increase with increasing latitude, mostly in populations with distribution in Andes range (Bekessy et al., 2002), or have found a high level of genetic variability with an important proportion of the total genetic diversity among populations (Ruiz et al., 2007).

However, one challenge in interpreting data from these studies is that they were based on a set of molecular markers which all have limitations in terms of understanding genome-wide levels of genetic variation and differentiation. For instance, plastid DNA markers are useful for detecting phylogeographic history, but their slow mutation rates, their uniparentally inherited nature and linkage of all plastid markers restricts insights into general population structure. Likewise, although isozymes are bi-parentally inherited and include multiple independent loci, they are fundamentally conserved loci, potentially subjected to selection, and limited to a small number of loci that can be easily assayed (typically ten or less). Finally, several studies on southern South American conifers have involved Randomly Amplified Polymorphic DNA (RAPDs). This approach, although providing data from multiple nuclear loci has been largely discredited based on lack of reproducibility of the fragment profiles the technique produces (Penner et al., 1993).

3.1.5 Objectives

To gain additional insights into the population genetic structure of Chilean conifers it is desirable to access large numbers of unlinked nuclear DNA markers. RAD-seq is applicable to species with no pre-exiting genomic data, and can provide access to large numbers of SNPs which can be used to assess genetic diversity and genetic differentiation.

In this study, I have used RAD sequencing to investigate the level of diversity and population genetic differentiation in four endemic conifer species from southern South America. Specifically, I assess (a) general levels of population genetic diversity and population structure, (b) whether there is detected differentiation associated with geographical location (e.g. the Andes vs Coastal range), and (c) whether population genetic variation and levels of differentiation are evenly or unevenly distributed throughout the species ranges.

3.2 Methods

3.2.1 Species

Samples were collected from four South American endemic conifer species with a restricted area of distribution and of conservation concern. Three Podocarpaceae species; *Saxegothaea conspicua* Lindley, *Prumnopitys andina* (Poepp. ex Endl) de Laubenfels, *Podocarpus salignus* D. Don and one member of the Cupressaceae family; *Fitzroya cupressoides*.

Saxegothaea conspicua is diploid and monoecious species with chromosome number 2n= 24 (Zonneveld, 2012). It grows in Chile and Argentina, with most (90%) populations location in the temperate forest of Chile (35-46° S). In the northern part of its distribution in Chile, *S. conspicua* grows between 800 and 900 m above sea level. *Saxegothaea conspicua* occurs in both the Coastal range and the Andes, and there are a marked temperature and aridity gradient across this part of its range. However, the annual mean temperature is about 9 ℃ with a mean annual precipitation of 1690 mm (Biffin et al., 2012). Though, temperatures at the Andean range can drop to -12 ℃ with annual precipitation of about 4000-5000 mm (Bannister and Neuner, 2001).

Prumnopitys andina is a dioecious (rarely monoecious) species with chromosome number 2n= 38 (determination from a cultivated plant) (Hair and Beuzenberg, 1958). It is uncertain if the species is diploid. However, the closer taxa in its phylogeny show similar numbers of chromosomes, therefore is also considered most likely to be a diploid species (Zonneveld, 2012). Its distribution is mostly in the Andean mountain from 35-39° S. It also has a single location on the eastern slopes of the Coastal range in La Araucanía region and another location in the Central valley, in Malleco (both at 38° S). Its altitudinal range varies between 300-1330 m above sea level. *Prumnopitys andina* occurs in locations with a Mediterranean climate with annual precipitation ranging between 2000 mm and 3000 mm (Donoso Zegers, 2006). However, the species prefers temperature between 0-9 ℃ and precipitation of about 1700 mm. The species is highly resistant to lower temperatures and occurs in locations where the winter temperatures drop between -6 ℃ to -12 ℃ (Bannister and Neuner, 2001).

Podocarpus salignus is also dioecious with a chromosome number first recorded as 2n= 38 (Hair and Beuzenberg, 1958) which was later recorded as 2n= 40 (Hair, 1966). It is uncertain if the species is diploid, the closer taxa in its phylogeny show similar numbers of chromosomes. However, other *Podocarpus* species show a lower number of chromosomes (ranging 2n= 20-24) (Hair, 1966). *Podocarpus salignus* is endemic species from Chile, distributed along both the Andean and Coastal mountain ranges of Chile from 35-40° S (Donoso Zegers, 2006). The altitudinal range of the species is from sea level up to 1000 m. The greatest differences in environmental conditions are between the warmer and drier Coastal range and the cooler and wetter Andes. Within its range of distribution, the mean annual temperature varies between 1-10°C and with annual precipitation of about 1700 mm (Biffin et al., 2012). ***Fitzroya cupressoides*** is a dioecious and tetraploid species with chromosome number 2n=4x= 44 (Zonneveld, 2012). *Fitzroya* is a very long-lived dioecious species, with some individuals age estimated at up to 3620 years (Lara and Villalba, 1993). *Fitzroya* occurs in a range of different environments in the wild. The species usually occur in a variety of vegetation types at altitudes between 100 and 1200 m above sea level, with most of the forests occurring between 500 and 800 m above sea level. It occurs in the Andes, the Central valley in Chile, and in the Coastal mountain range in Chile. More specifically, *Fitzroya* occupies a large climatic range, with annual precipitation between 2000 and 4000 mm (Allnut et al. 2001, Veblen et al. 1976).

3.2.2 Seed dispersal mechanism of the conifer species

Overall, the seed dispersal mechanisms slightly differs between these conifer species. *Saxegothaea conspicua* produce a female cone that usually falls by gravity. No seed dispersers are known. It may be possible that rodents disperse female cones a short distance. *Prumnopitys andina* produce a green acid edible fruit with a single seed. The fruit usually falls by gravity and can also be dispersed by rodents. Farm animals usually eat this fruit and in some cases (cows) promote the dispersal and germination of the seed (because of the natural scarification process). The seed is very hard to germinate, needing long cold exposure (7-9 month, or usually two winters). *Podocarpus salignus* produces the typical false fruit of the Podocarpaceae species. It also falls by gravity

but can be dispersed before falling by birds that are attracted by its "fruit". *Fitzroya cupressoides* produce a female cone with many seeds inside. The cones usually open before falling, and the seeds are dispersed by wind. However, the species only produce seeds every 5-7 years and only about 10-30% of the seeds by cone develop embryos (i.e. very low-rate of seed viability).

3.2.3 Sampling

A total of 113 needle samples were collected throughout the entire natural distribution of the four conifers in Chile, and locations were chosen to include areas from the Andes, Coastal range and the Central valley. Between 7-9 populations were included per species (Figure 3.1) and between 2-5 samples per population. Each individual was randomly selected with at least 50-meters distance from each other to avoid sampling closely related individuals, and leaf tissue (needles) was stored in silica gel after collection.

3.2.4 DNA extraction, RAD-Seq library construction and sequencing

DNA was extracted (from leaf tissue) using the DNeasy^{TM} Plant Minikit, (QIAGEN, Netherlands), following the manufacture's protocol. Library construction for RAD-seq (using the enzyme PstI), for all four data sets, was performed following similar methods to Baird et al. (2008), and it was conducted by an external company; Floragenex, USA. who provided the raw data for further analyses. A full description of the laboratory protocols is given in the Material and Methods of Chapter 2.2.2.

3.2.5 Data analysis

Raw reads were processed with Stacks version 2.4. First, *process_radtags* was executed to demultiplex reads that have suitable barcode sequence and RAD cut-site. A raw Phred quality score filter of 10 was applied (removing any read below 10, details in Chapter 2). Due to the lack of a reference genome in each species, *de novo* assembly was performed in each data set. The main parameters that control the *de novo* assembly in Stacks (m, M, n) were optimised following the protocol of Paris et al. (2017) and Catchen et al. (2011) in order to reduce biases and problems of under or over-merging reads (Catchen et al., 2013) (details in Chapter 2). The *de novo* assemblies were completed

with the following optimised parameters; *S. conspicua* and *P. salignus*: m5 M4 n4; *Pr. andina*: m4 M3 n3; *F. cupressoides*: m4 M2 n2. More details in Material and Methods, Chapter 2.2.5.

3.2.6 Data filtering and calculation of population differentiation

Genotyping was carried out across all species-samples using the *populations* pipeline of Stacks. Applying the R filter, I recovered loci that were only present in at least in 60% of the total number of individuals in each data set (R60) (Chapter 2). Next, using Tassel version 5.0 (Bradbury et al., 2007) the level of missing data was calculated in each data set. Individuals with a high level of missing data (>50%) were removed (13 samples in total; 3 samples in *S. conspicua*, *Pr. andina* and *F. cupressoides* and 4 samples in *P. salignus*). With the remaining samples (see in Table 3.1), the *populations* pipeline of Stacks was run again, also recovering loci that were only present in at least in 60% of the total number of individuals in each data set (R60). A minor allele frequency (MAF) of 0.02 and maximum observed heterozygosity of 0.5 was also applied. This to discard rare alleles and minimise potential paralogs, respectively. Using the function write-single-snp from the *population* pipeline in Stacks, only unliked SNPs were recovered. Finally, only loci that were present in at least in one sample per population were selected. This is to ensure that genotype information is present in all populations and minimise potential biases due to missing data.

3.2.7 Genetic diversity

The level of nucleotide diversity (pi-value-per site) in each species-population was estimated using the *populations* pipeline of Stacks. The pi calculation is defined in Stacks by equation 2 from Hohenlohe et al. (2010).

3.2.8 Population structure

Using diveRsity package for R version 3.6, global F_{ST} and pairwise F_{ST} was calculated with 1000 bootstrap replicates for global F_{ST} and 100 bootstraps replicates for pairwise F_{ST} (F_{ST} is based on Weir and Cockerham 1984). The pattern of population differentiation for each species was examined in the form of isolation by distance (IBD, Slatkin 1993 and Slatkin 1987) for all populations for a given species by correlating $F_{ST}/(1\text{-}F_{ST})$ with geographical distance (Km) which was simply estimated as a linear distance between sampling localities. A mantel test with 999 permutations was used to evaluate if there is a significant correlation, as implemented in GenAlEx. To assess the relationships between populations, a Discriminant Analysis of Principal Components (DAPC, Jombart et al. 2010) was performed using the R package adegenet (Jombart 2008; Jombart and Ahmed 2011). To establish the optimal number of principal components for the DAPC analysis, cross-validation was performed with 100 replicates. To further explore genetic structure, clustering was undertaken using STRUCTURE software version 2.3.4 (Pritchard et al., 2000). Structure was run 10 times for each K-value from 1-9, depending on the maximum number of populations in each data set. These analyses used a burn-in of 10,000 followed by 100,000 MCMC replicates using the admixture and correlated allele frequencies options. To detect the optimal K value, the Evanno (Evanno et al., 2005) method was used in the Clustering Markov Packager program (CLUMPAK) (Kopelman et al., 2015).

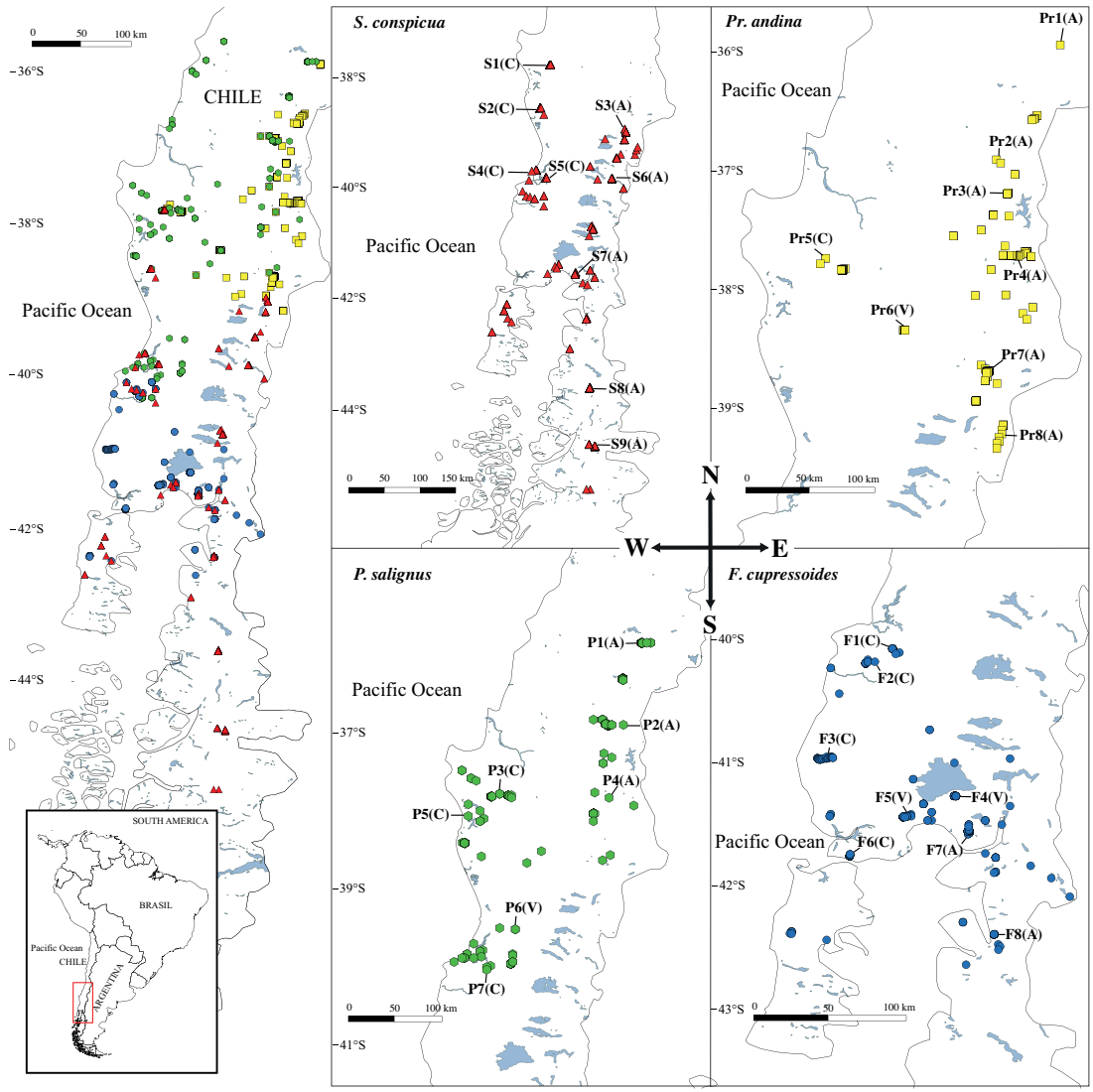


Figure 3.1: Map showing the entire natural distribution range, and populations sampled of each conifer species in Chile. On the left the distribution of each conifer species in Chile (in red *S. conspicua*, in yellow *Pr. andina*, in green *P. salignus* and in blue *F. cupressoides*). On the right, distribution of each conifer specie indicating the populations included during the investigation (S1-S9 for *S. conspicua*, Pr1-Pr8 for *Pr. andina*, P1-P7 for *P. salignus* and F1-F8 *F. cupressoides*). In parenthesis is indicated the region of each population; (A) Andean mountain range; (C) Coastal mountain range and (V) Central valley. The occurrence points for the distribution of the four conifer species were obtained from data provided by the Royal Botanic Garden of Edinburgh and personal information.

Table 3.1: Name, location, the total number of samples, altitude and estimated population size of each conifer species. Samples that were removed due to high level of missing data are not listed. Populations are ordered by latitude.

No	Population name	Region	Total sample	Latitude (S)	Longitude (W)	Altitude (m)	Approx. size No trees
			Saxegothaea conspicua				
S1	Nahuelbuta	Coastal	5	37°48'42"	73°03'54"	979	<150
S2	Villa las Araucarias	Coastal	2	38°34'24"	73°13'44"	609	>2000
S3	Nasampulli	Andes	2	39°01'00"	71°41'14"	1229	<500
S4	Oncol	Coastal	2	39°41'53"	73°18'06"	485	<500
S5	Llancahue	Coastal	2	39°50'15"	73°08'00"	292	<400
S6	Huilo Huilo	Andes	3	39°51'14"	71°57'06"	485	<350
S7	Lenca	Andes	2	41°34"17"	72°35'57"	697	<500
S8	Río Futa	Andes	3	43°36'08"	72°20"21"	140	>1000
S9	Rio Cisne	Andes	5	44°41'04"	72°14'51"	179	>1000
			Prumnopitys andina				
Pr1	Corral de Salas	Andes	5	35°52'58"	70°59'22"	1022	<70
Pr2	Los Lleuques, Antuco	Andes	3	36°51'40"	71°36'16"	957	<200
Pr3	Laja	Andes	3	37°22'25"	71°33'53"	833	<100
Pr4	Trapa-Trapa	Andes	3	37°40'13"	72°01'12"	788	<700
Pr5	Nahuelbuta	Coast	5	37°49'30"	72°48'29"	871	<100
Pr6	Pua-Santa Lucia	Central valley	2	38°20'26"	72°19'09"	395	<400
Pr7	Conguillio	Andes	3	38°49'36"	71°39'44"	542	<600
Pr8	Reigolil	Andes	3	39°08'08"	71°28'58"	826	>2000
			Podocarpus salignus				
P1	Hornillos	Andes	5	35°51'03"	71°10'13"	791	<900
P2	Los Lleuques, Antuco	Andes	3	36°21'37"	71°41'20"	747	>2000
P3	Nahuelbuta	Coast	3	37°49'10.6"	73°05'46"	936	<200
P4	Los Guindos	Andes	3	38°02'07"	71°46'50"	602	<400
P5	Tirua	Coast	5	38°24'43"	73°26'16"	457	<300
P6	Reumen	Central valley	2	38°57'31"	72°50'29"	69	<600
P7	Llancacura	Coast	5	40°17'15"	73°20'44"	250	<200
			Fitzroya cupressoides				
F1	Anchile	Coast	3	40°04'42"	73°12'55"	439	<100
F2	Alerce Costero	Coast	2	40°11'38"	73°26'03"	924	>1500
F3	C. pelada	Coast	2	40°57'41"	73°49'28"	662	>2000
F4	Río pescado	Central valley	3	41°16'30"	72°42'19"	271	<500
F5	F. nuñez	Central valley	3	41°25'01"	73°08'00"	70	<200
F6	Astillero	Coast	3	41°45'18"	73°33'53"	43	<700
F7	Lenca	Andes	2	42°23'16"	72°23'39"	893	>2000
F8	Huinay	Andes	3	42°23'28"	72°23'55"	867	>2000

3.3 Results

3.3.1 Genetic diversity

Between 853 and 4915 single nucleotide polymorphisms were generated for the four conifer species. Most SNPs were retained in *Pr. andina* (4915 SNPs), followed by *P. salignus* with 4212 SNPs and *S. conspicua* with 3137 SNPs. The fewest SNPs were found in *F. cupressoides*, with a total of 853 SNPs (Table 3.3). The level of nucleotide diversity within populations ranged between pi= 0.11 and 0.27%. The highest diversity was found in *P. salignus* populations, ranging between pi= 0.16 and 0.27%. In contrast the lowest nucleotide diversity was found in *F. cupressoides* ranging between pi= 0.11 and 0.17% (see Table 3.2).

Table 3.2: Mean sample size per locus and nucleotide diversity (pi value) for each population in each conifer species

S. conspicua			*Pr. andina*		
Population	Mean sample per locus	pi value (%)	Population	Mean sample per locus	pi value (%)
S1	3.766	0.22	Pr1	3.725	0.22
S2	1.56	0.16	Pr2	1.991	0.17
S3	1.364	0.12	Pr3	1.907	0.16
S4	1.689	0.23	Pr4	2.258	0.20
S5	1.459	0.14	Pr5	3.754	0.22
S6	2.01	0.17	Pr6	1.42	0.16
S7	1.482	0.14	Pr7	2.064	0.18
S8	1.911	0.15	Pr8	2.397	0.22
S9	3.494	0.23			
P. salignus			*F. cupressoides*		
P1	3.636	0.27	F1	1.382	0.11
P2	2.014	0.20	F2	2.263	0.17
P3	2.034	0.20	F3	2.065	0.16
P4	2.106	0.22	F4	1.956	0.17
P5	3.527	0.27	F5	2.019	0.15
P6	1.446	0.17	F6	1.577	0.15
P7	3.647	0.27	F7	2.081	0.15
			F8	1.504	0.12

3.3.2 Population structure

Low levels of genetic differentiation were observed in all four species with global measures of population differentiation ranging from F_{ST}= 0.062 in *S. conspicua*, F_{ST}= 0.060 in *Pr. andina*, F_{ST}= 0.045 in *P. salignus* and F_{ST}= 0.017 in *F. cupressoides*. In *Fitzroya* the estimates of F_{ST} were not significantly different from zero (p>0.05, see Table 3.3).

The genetic differentiation between each pair of populations was also low with a pairwise F_{ST} ranging from -0.008 to 0.175 (see Table 3.4). None of the pairwise comparisons between populations showed significant differences in any species. Moreover, the Mantel test for the four conifer species showed that there was no significant relationship between genetic distances (F_{ST}/(1-F_{ST})) and geographical distances (see Figure 3.2).

When the four conifer species were analysed with DAPC, there was a weak signal of population clustering in three of the four species (*S. conspicua*, *Pr. andina* and *F. cupressoides*). The species with the widest distribution (*S. conspicua*), showed a slight division between the northern and the southern Andes populations and also with some coastal populations. The other species with a smaller distribution (*Pr. andina*, *P. salignus* and *F. cupressoides*) showed less clear division between populations and regions such as *P. salignus* that retained only one discriminant function, grouping all populations as a single cluster. The DAPC, in *F. cupressoides* tended to divide the Coastal range populations and the Andes and Central valley populations (both Andes and valley grouped). However, the general pattern for all four conifer species showed that individuals from different regions and populations also cluster together, with no evidence for clear differentiation between them.

When the STRUCTURE analysis was run and explored using the Evanno method, an "optimal" K= 6 was identified for *S. conspicua* and *P. salignus*, K= 2 for *Pr. andina* and K= 5 for *F. cupressoides*. However, this clustering did not correspond to any grouping of individuals by their source population, or of proximal populations, and it is important to note that a small number of individuals-samples may generate biases to find the "optimal" K using this method (Waples and Gaggiotti, 2006). Moreover,

the Evanno method does not allow the assessment of $K = 1$ as a potential solution (Namroud et al., 2008).

Table 3.3: Global F_{ST} for each conifer species. Calculation made with a 1000 bootstrap replicates (BC 95% confident interval, c.i).

Species	F_{ST}	c.i. lower	c.i upper	no. SNPs
S. conspicua	0.062	0.051	0.073	3137
Pr. andina	0.060	0.053	0.068	4915
P. salignus	0.045	0.039	0.051	4212
F. cupressoides	0.017	-0.011	0.043	853

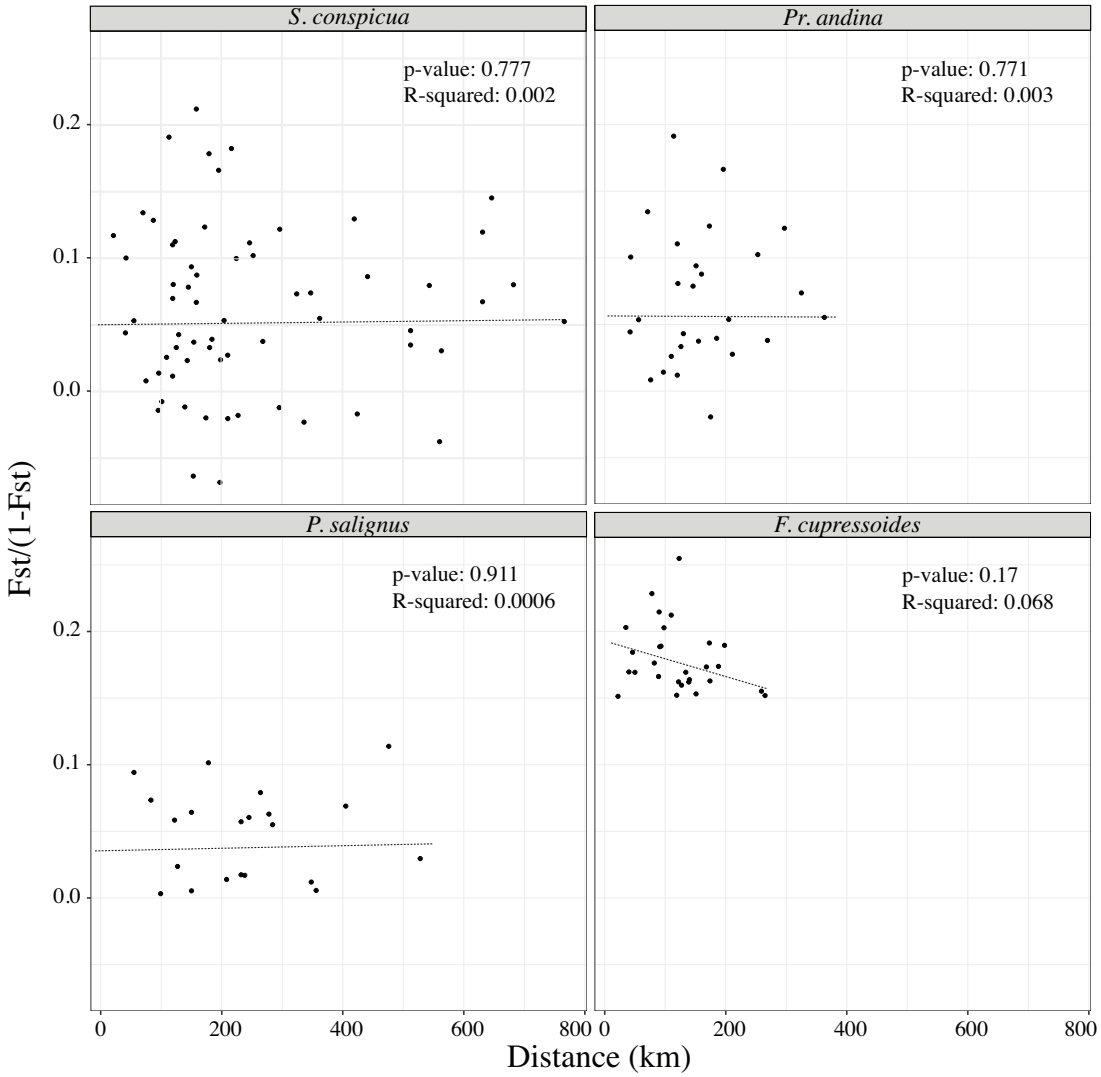


Figure 3.2: Isolation by distance of each conifer species, showing non-significant correlation.

Table 3.4: Pairwise F_{ST} population comparisons for each conifer species. Calculations made with a 100 bootstrap replicates. None of the species-populations showed significant genetic differences, P-value >0.05

S. conspicua

	S1	S2	S3	S4	S5	S6	S7	S8
S1								
S2	0.114							
S3	0.151	0.022						
S4	-0.021	0.101	0.175					
S5	0.091	-0.012	-0.068	0.105				
S6	0.100	0.032	-0.015	0.062	-0.008			
S7	0.114	-0.024	-0.013	0.154	-0.074	0.023		
S8	0.127	0.039	0.033	0.079	-0.018	0.043	-0.019	
S9	0.050	0.074	0.107	-0.040	0.073	0.063	0.069	0.065

Pr. andina

	Pr1	Pr2	Pr3	Pr4	Pr5	Pr6	Pr7
Pr1							
Pr2	0.075						
Pr3	0.110	0.051					
Pr4	0.051	0.014	0.043				
Pr5	0.037	0.086	0.100	0.041			
Pr6	0.109	-0.020	0.032	0.025	0.119		
Pr7	0.069	0.027	0.036	0.012	0.073	0.008	
Pr8	0.052	0.093	0.143	0.081	0.038	0.161	0.091

P. salignus

	P1	P2	P3	P4	P5	P6
P1						
P2	0.055					
P3	0.073	0.005				
P4	0.057	0.023	0.003			
P5	0.012	0.054	0.068	0.060		
P6	0.102	0.006	0.017	0.017	0.092	
P7	0.029	0.0654	0.065	0.052	0.014	0.086

F. curpressoides

	F1	F2	F3	F4	F5	F6	F7
F1							
F2	-0.047						
F3	0.038	-0.024					
F4	0.027	0.012	0.030				
F5	0.059	0.025	0.050	0.003			
F6	0.024	-0.039	0.013	-0.031	-0.017		
F7	0.063	-0.012	-0.020	0.019	0.025	0.015	
F8	0.054	-0.017	-0.029	0.033	0.017	-0.021	0.005

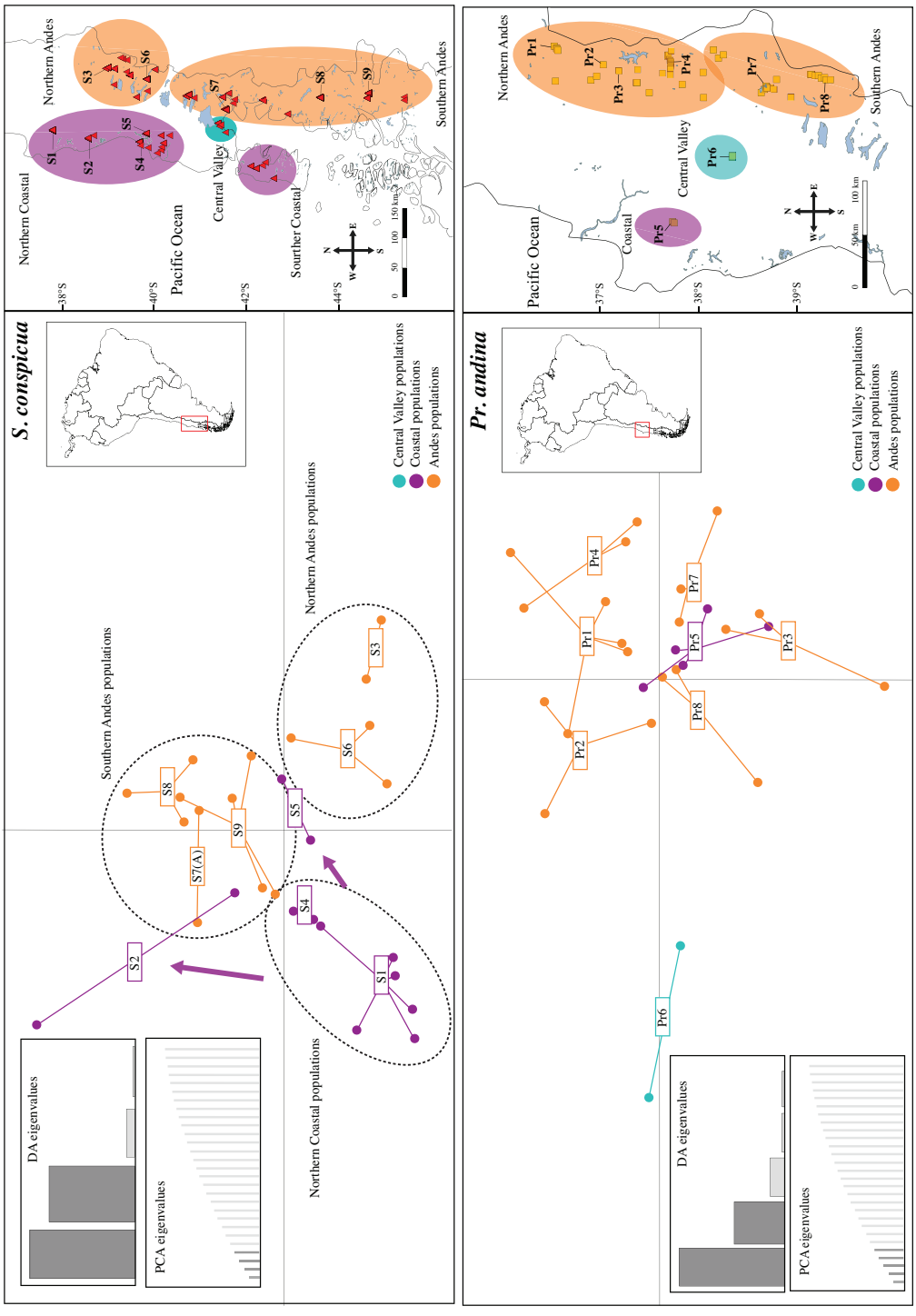


Figure 3.3: Discriminant Principal Component Analysis (DAPC) plot showing potential population clusters by region (Andes, Coastal and Central valley) for *S. conspicua* (top) and *Pr. andina* (bottom).

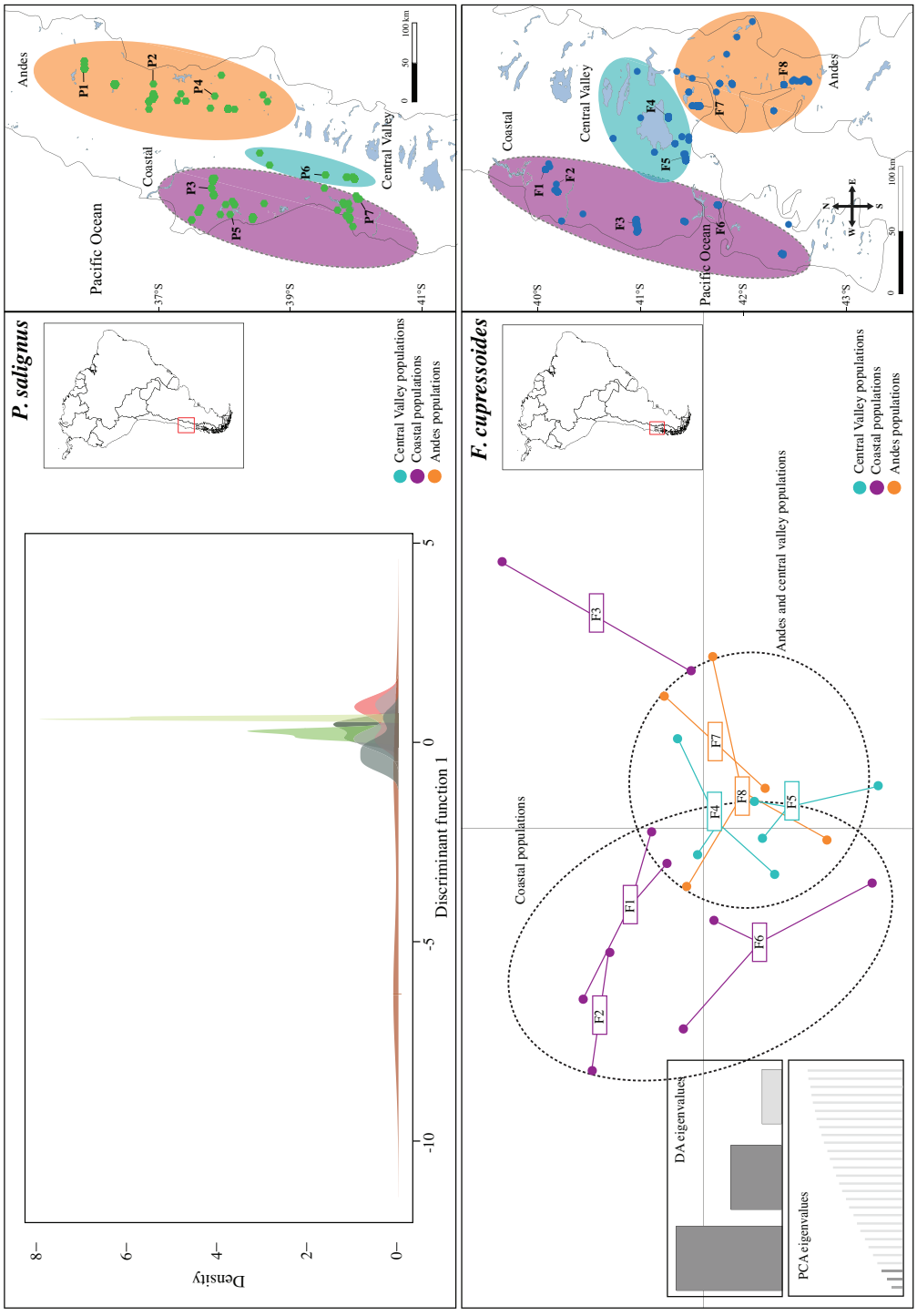


Figure 3.4: Discriminant Principal Component Analysis (DAPC) plot showing potential population clusters by region (Andes, Coastal and Central valley) for *P. saligna* (top) and *F. cupressides* (bottom).

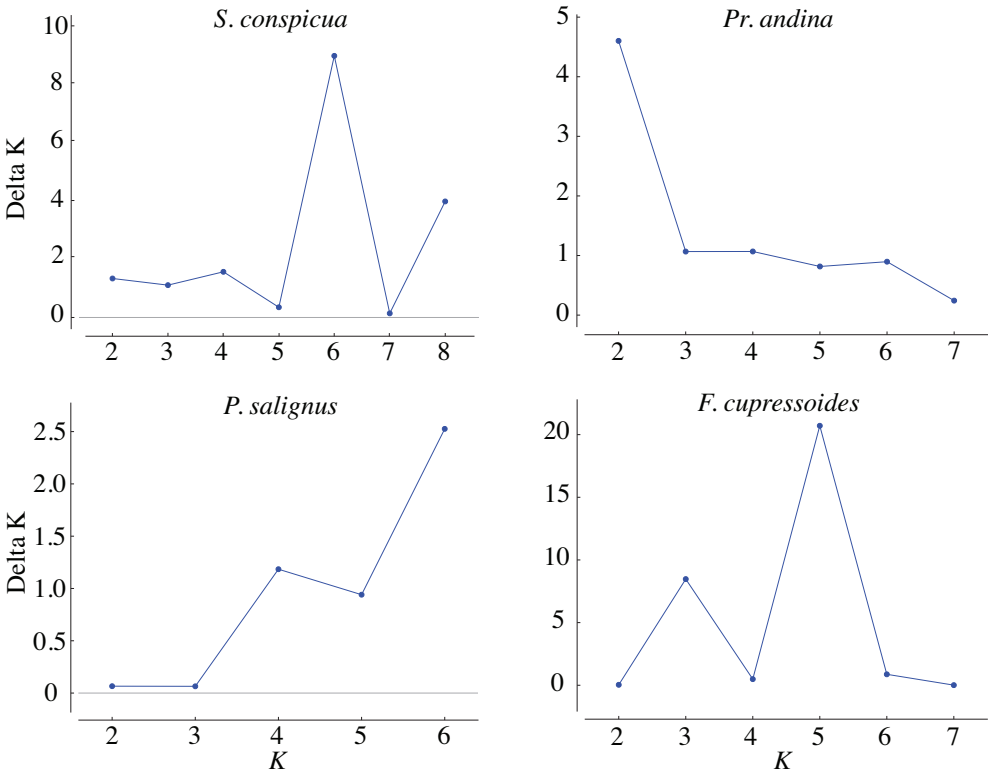


Figure 3.5: Probability of the "optimal" K according to the Evanno method for all four species.

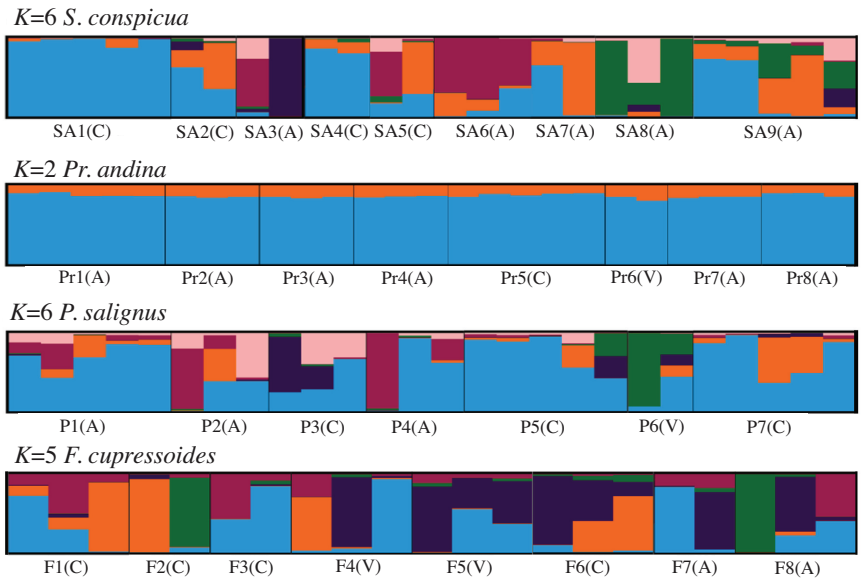


Figure 3.6: Major groups identified for all species-populations, illustrating clumpak's label-matching across K values. Only the optimal K identified by the Evanno method for each species is illustrated. In parentheses is indicated the region of each species-population; Andes (A), Coastal (C) and Central valley (V).

3.4 Discussion

This study of RAD-seq variation in four endemic southern South American conifer species detected broadly similar levels of genetic diversity within each species, and a consistent signature of low levels of population differentiation.

Genetic diversity

This study identified between 3000 and 5000 SNPs in the Podocarpaceae species and about 900 in *F. cupressoides*. The average nucleotide diversities in *S. conspicua*, *Pr. andina*, *P. salignus* and *F. cupressodoides* ranged between pi= 0.11-0.27%. These values are slightly lower than those in other conifers from the north hemisphere such as; *Cryptomeria japonica* (L.f.) D.Don (Kado et al., 2003), *Abies alba* Mill., *Larix decidua* Mill., *Pinus cembra*, *Pinus mugo* Turra (Mosca et al., 2012), *Pinus thunbergii* Parl, *Pinus densiflora* Siebold & Zucc (Suharyanto and Shiraishi, 2011), *Pinus menziesii var menziesii* (Mirbel) Franco, (Eckert et al., 2009), *Pinus sylvestris* L. (García-Gil et al., 2003) and *Pinus taeda* L. (Brown et al., 2004), ranging between pi= 0.2-0.8%. There are no comparable studies of southern South American conifers with sequence data from the nuclear genome. There are studies such as Quiroga et al. (2012) that report sequence diversity for a South American Podocarpaceae, *Podocarpus parlatorei* Pilg., from the plastid genome (pi=0.024%). However, these are not suitable for comparison, as the slower mutation rate, and reduced effective population size of the uniparentally inherited plastid genome confound comparison (the effective population size of the plastid genome in dioecious conifers is 1/4 that of the nuclear genome (Ennos et al., 1999).

Population structure

Genetic differentiation was low in all four species, as measured by global F_{ST} (range: 0.017-0.062) or indicated by the lack of population or regional clustering in the DAPC and STRUCTURE analyses. To date, very little other modern genetic work has been done on the same species. Using RAPD markers, Allnutt et al. (1999) and Allnut et al. (2001) reported mean Phi_{ST}= 0.143 in *F. cupressoides* and mean Phi_{ST}= 0.069 in *P. salignus*. In contrast, an unpublished PhD study using isozyme markers observed a higher level of genetic differentiation in *P. salignus* with F_{ST}= 0.448 (Quiroga, 2009). The same study reported F_{ST}= 0.082 in *S. conspicua* and F_{ST}= 0.140 in *Pr. andina*, the latter with an F_{ST} value not statistically different from zero (p-value>0.05). However, another unpublished research using cpDNA markers by Martinez Araneda (2011) observed a G_{ST} = 0.184 in *Pr. andina* (p-value>0.05, significantly different from zero).

Overall, the levels of genetic differentiation between populations (F_{ST} pairwise values) found in this study for *F. cupressoides*, *Pr. andina* and *P. salignus* differ slightly from past research on the species, with this study showing lower levels of population differentiation. For example, this investigation revealed low genetic differentiation between populations in all species with none of the pairwise F_{ST} values statistically different from zero, and a non-significant relationship between genetic distances and geographical distances in all four species (Figure 3.2). Allnutt et al. (1999); Allnut et al. (2001) and Martinez Araneda (2011) detected low but a significant genetic differentiation among populations for *F. cupressoides*, *Pr. andina* and *P. salignus* respectively. These differences could be associated with the different methods used. Population structure insights from RAD markers are clearly different from RAPD and cpDNA markers and direct comparisons are not easy. For example, RAPD provides data from multiple nuclear loci; however, the lack of reproducibility of these markers could generate biases. On the other hand, a major difference between this study and those undertaken previously, is that the current study has focused on getting more data per sample (hundreds of loci) from a small number of individuals, where the previous studies have sampled fewer loci from more individuals. In addition, in the case of *Fitzroya*, Allnutt et al. (1999) in their investigation they sampled broader geographic scope incorporating populations

from the eastern part of the Andes range in Argentina. Therefore the isolation between the Argentinian and Chilean populations (by the Andes range) might also be a factor involved in the levels of genetic differentiation observed between populations in this species.

Table 3.5: Comparison of level of genetic differentiation found in different investigations (including this study). * values not statistically different from zero.

Species	Genome	Marker type	Measure of differentiation	Estimated differentiation	Reference
S. conspicua	Nuclear	RAD	F_{ST}	0.062	This investigation
Pr. andina	Nuclear	RAD	F_{ST}	0.060	This investigation
P. salignus	Nuclear	RAD	F_{ST}	0.045	This investigation
F. cupressoides	Nuclear	RAD	F_{ST}	0.017*	This investigation
Level of genetic differentiation found in other investigations					
S. conspicua		Isozyme	F_{ST}	0.082	Quiroga, 2009 (PhD Thesis)
Pr. andina		Isozyme	F_{ST}	0.140*	Quiroga, 2009 (PhD Thesis)
P. salignus		Isozyme	F_{ST}	0.448	Quiroga, 2009 (PhD Thesis)
P. salignus	Nuclear	RAPD	Phi_{ST} (mean pairwise)	0.069	Allnut et al. (2001)
F. cupressoides	Nuclear	RAPD	Phi_{ST} (mean pairwise)	0.143	Allnutt et al. (1999)
Pr. andina	Plastid	CpDNA	G_{ST}	0.184	Martinez Araneda (2011) (PhD Thesis)

Conifers are typically thought to show low population genetic differentiation, even among geographically distant populations (O'Connell et al., 2007). One factor that is postulated as important here is very effective wind-borne pollen dispersal, which may contribute to maintaining common alleles over large geographical distances (Prunier et al., 2016). Indeed, Hamrick et al. (1992) in their review of 121 gymnosperms species revealed little genetic differentiation among populations, with an overall mean G_{ST}= 0.073 (based on allozyme analyses). Based on SNP arrays, only a few direct studies of population genetics have been undertaken so far, e.g. Namroud et al. (2008), which also identified very low level of genetic differentiation in the American white spruce (*Picea glauca* (Moench) Voss, F_{ST}= 0.006). However, all these observations are dominated by northern hemisphere boreal conifers that frequently occur in large continuous forest blocks. Thus the high level of connectivity found in this investigation between populations is somewhat unexpected, as the four species have a discontinuous distribution in Chile, with many populations separated by marked topographical barriers (discussed later).

Potential non-biological reasons for low population structure

Before considering biological interpretations of low population structure, it is essential to rule out technical artefacts. For instance, it is possible that population differentiation is present, but that the samples sizes were too small to detect it, or that missing data or treatment of paralogous loci as orthologues results obscured this population differentiation. In the case of small sample sizes (between 2-5 per pop), I am confident this does not lead to a problem. This is mainly because RAD molecular markers provide a huge amount of data (SNPs) with the small sample size of individuals being compensated to some extent by the large sample size of loci from each individual. It is also noteworthy that in individual-level analyses – no geographic or population-specific clustering is evident. This is shown in the DAPC analysis (and was also evident in basic PCA analysis and Neighbour-joining trees – results not shown). Clearly, additional sampling of more individuals per population would be desirable, but this repeated lack of geographical structure in population and individual-based analyses suggest that the finding of low differentiation is robust to sample size effects.

In the case of missing data, I am also confident this does not lead to a problem. A sensitivity analysis was conducted previously in Chapter 2. This analysis concluded that the level of missing data did not impact on the level of genetic structure in each of the conifer species as measured by F_{ST}. I have further examined whether this impacts on individual-based analyses and considered whether changing the missing data threshold from 40% and 20% (R60 and R80) would impact DAPC clustering. Although there are impacts on the distribution of points in the plot, there is no material difference in population-specific clustering or clustering together of individuals from adjacent populations or regions (illustrated in Appendix B.1 for *S. conspicua*).

Regarding the paralogy issue, this is more difficult to exclude as by definition – undetected paralogy would remain unknown. However, if there was an excess of paralogous loci in the dataset, we would expect to see highly inflated observed heterozygosity, if different loci are mistakenly treated as different alleles. No such heterozygosity excess was observed. It is also noteworthy that F_{ST} was robust to different assembly parameters (default values and optimised values, see Chapter 2) suggesting that there

is considerable stability in the data and an overriding signature of low population differentiation.

Factors that might impact population structure in Chilean conifers

Most of the extant populations of Chilean conifers are restricted to the slopes of the Andean cordillera and to a lesser extent in the Coastal range. This is mostly due to the impact of human activities, which have reduced a large proportion of the Central valley and Coastal range forest (Vergara et al., 2013). This range fragmentation coupled with natural barriers to gene flow stemming from the complex topography of Chile might be expected to lead to high levels of population differentiation. However, low-levels of genetic differentiation and weak geographical structure in clustering analyses were detected. This implies that the four species have not (yet) been affected by these geographical barriers and/or fragmentation. This might be associated with the intrinsically effective wind-borne pollen dispersal of these species. However, this alone seems an unsatisfactory explanation, given the scale of the barriers and the distances between extant populations.

The sheer longevity of individuals of these species may be the critical factor in determining the observed levels of population differentiation. The four conifers involved in this research are long-lived trees, reaching from between 200-700 years for the Podocarpaceae species to over 3000 years in *F. cupressoides* (Lara and Villalba, 1993). Indeed, one of the Podocarpaceae species; *S. conspicua* shows a striking and widespread evolutionary adaptation; adventitious roots (not seen in another Podocarps species yet). These adventitious roots grow down inside the hollowed-out bark of old trees of *S. conspicua* that allow the individuals to anchor itself to the ground with the consequence of perpetuation *in situ*, extending their longevity. This is a process that might be repeated in individuals from generation to generation; thus, this species could be listed as one of the most long-lived trees in the world (Cano et al., 2014).

The major anthropogenic impacts of these species have been in the last 500 years for *F. cupressoides* and about 100 years for the Podocarpaceae species. Likewise, before the 1970s the exploitation and native forest reduction, including the *Fitzroya* forests, were caused by the Spanish colonists and later by agricultural activities (Gardner et al.

2006, Robles Ortiz 2003). Over the 16-17th centuries, the first settlers burned large areas of forests (Gardner et al., 2006). They also began to over-exploited *Fitzroya* for its timber, an activity that became one of the primary economic resources in the south of Chile (Torrejón et al., 2011). The expansion of the agricultural activities began in the 19th century and mostly affected the Central valley of the central southern part of Chile (Robles Ortiz, 2003). After the 1970s the most extensive forest reduction (at least the Coastal range and Central valley) was caused by the expansion of the forestry industry (*Pinus radiata* and *Eucalyptus*) that is still affecting a significant part of the native forests of Chile (Salas et al., 2016).

Over deeper timescales, species history might also have shaped the current patterns of genetic structure in these conifers. Palynological and glaciological evidences suggests that the last glaciation maximum (LGM) affected the distribution of the southern Andean forest, reducing its size (compared to the current distribution) due to extensive ice coverage (Allnutt et al. 2003, Villagrán 1991). Two main hypotheses have been proposed to explain the effect of the LGM on the existing tree species distributions in southern South America. The first is that tree species survived in multiples refugia and the second hypothesis is that they recolonised from a single refugium (Markgraf et al., 1995). Using isozyme markers Premoli et al. (2000b) assessed these hypotheses in *F. cupressoides*. They suggested that if the current populations originated from only one refugium, low-levels of genetic differentiation should be detected across the range of *F. cupressoides* including between populations in the Andes and the Coastal mountains. In contrast, multiple refugia could lead to a higher degree of genetic divergence between populations. Results by Premoli et al. (2000b) detected a degree of genetic variation between eastern and western populations (based mostly on DAPC analysis), leading them to suggest that the current *F. cupressoides* distribution would have originated from multiples refugia. Several other investigations have also suggested the existence of multiple small refugia for different species in the region, based on RAPD or isozyme nuclear markers (e.g. Marchelli and Gallo 2004, Bekessy et al. 2002, Pastorino and Gallo 2002) and plastid markers (e.g. Pastorino et al. 2009, Paula and Leonardo 2006, Marchelli et al. 1998). In contrast, this investigation, based on SNPs arrays (RAD-seq) showed no signal of genetic structure in any of the four conifers. In this context, our results would

suggest that might be reasonable to not rule out the hypothesis of a forest expansion from a single refugium. Another related hypothesis, from Marchelli et al. (2010), who observed no genetic differentiation between the *A. araucana* (endemic conifer species from the south of Chile) populations, suggested that the pre-Pleistocene distribution of the species could have been only partially reduced and it could be possible for there to have been multiple refugia without observing strong genetic differentiation among extant populations.

Conclusion

This work represents the first population genetics analysis using restriction site associated DNA markers on conifers from southern South America (*S. conspicua*, *Pr. andina*, *P. salignus* and *F. cupressoides*). This investigation demonstrated the ability to use RAD sequencing on large genome conifer species to investigate population genetic structure. The results indicated a similar level of nucleotide diversity and no genetic structure in all four conifer species. These low-levels of genetic differentiation are similar to the levels of genetic differentiation found in boreal conifers, which usually show much more extensive and continuous distribution than the southern hemisphere conifers. A key factor, potentially explaining this finding is the extreme longevity of the Chilean conifers, which may be important in retarding genetic drift despite their more isolated and fragmented ranges.

Bibliography

Allnut TR, Courtis JR, Gardner M, Newton AC. 2001. Genetic variation in wild Chilean and cultivated British populations of *Podocarpus salignus* D. DON (Podocarpaceae). *Edinburgh Journal of Botany*. 58:459–473.

Allnutt TR, Newton AC, Lara A, Premoli A, Armesto JJ, Vergara R, Gardner M. 1999. Genetic variation in *Fitzroya cupressoides* (alerce), a threatened South American conifer. *Molecular Ecology*. 8:975–987.

Allnutt TR, Newton AC, Premoli A, Lara A. 2003. Genetic variation in the threatened South American conifer *Pilgerodendron uviferum* (Cupressaceae), detected using RAPD markers. *Biological Conservation*. 114:245–253.

Awad L, Fady B, Khater C, Roig A, Cheddadi R. 2014. Genetic structure and diversity of the endangered fir tree of Lebanon (*Abies cilicica* Carr.): Implications for conservation. *PLoS ONE*. 9.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*. 3:1–7.

Bannister P, Neuner G. 2001. Frost resistance and the distribution of conifers, In: *Conifer Cold Hardiness*; Bigras, F.J., Colombo S.J., Eds.; Kluwer Academic Publisher: Boston, Ma, USA., pp. 3–21.

Bekessy SA, Allnutt TR, Premoli AC, Lara A, Ennos RA, Burgman MA, Cortes M, Newton AC. 2002. Genetic variation in the Monkey Puzzle tree (*Araucaria araucana* (Molina) K. Koch), detected using RAPD. *Heredity*. 88:243–249.

Biffin E, Brodribb TJ, Hill RS, Thomas P, Lowe AJ. 2012. Leaf evolution in southern hemisphere conifers tracks the angiosperm ecological radiation. *Proceedings of the Royal Society B: Biological Sciences*. 279:341–348.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*. 23:2633–2635.

Brown GR, Gill GP, Kuntz RJ, Langley CH, Neale DB. 2004. Nucleotide diversity and linkage disequilibrium in loblolly pine. *Proceedings of the National Academy of Sciences of the United States of America*. 101:15255–15260.

Cano MJ, Godoy R, Soto DP. 2014. Raíces adventicias de *Saxegothaea conspicua* Lindl. (Podocarpaceae) en su etapa de senescencia en bosques adultos. *Gayana - Botanica*. 71:171–174.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013. Stacks: an analysis tool set for population genomics. *Molecular Ecology*. 22:3124–3140.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. 2011. Stacks: Building and genotyping loci *de novo* from short-read sequences. *G3: Genes, Genomes, Genetics*. 1:171–182.

Colabella F, Gallo LA, Moreno AC, Marchelli P. 2014. Extensive pollen flow in a natural fragmented population of Patagonian cypress *Austrocedrus chilensis*. *Tree Genetics and Genomes*. 10:1519–1529.

Donoso Zegers C. 2006. Las especies arbóreas de los bosques templados de Chile y Argentina : autoecología. Valdivia: Maria Cuneo, primera edition.

Eckert AJ, Wegrzyn JL, Pande B, Jermstad KD, Lee JM, Liechty JD, Tearse BR, Krutovsky KV, Neale DB. 2009. Multilocus patterns of nucleotide diversity and divergence reveal positive selection at candidate genes related to cold hardiness in coastal Douglas fir (*Pseudotsuga menziesii var. menziesii*). *Genetics*. 183:289–298.

Eliades NGH, Gailing O, Leinemann L, Fady B, Finkeldey R. 2011. High genetic diversity and significant population structure in Cedrus brevifolia Henry, a narrow endemic Mediterranean tree from Cyprus. *Plant Systematics and Evolution*. 294:185–198.

Ennos RA, Langdon A, Hu XS, Sinclair WF. 1999. Using organelle markers to elucidate the history, ecology and evolution of plant populations. *Heredity*. 72:250–259.

Frankham R, Ballou JD, Briscoe DA. 2004. A Primer of Conservation Genetics. Cambridge: Cambridge University Press.

García-Gil MR, Mikkonen M, Savolainen O. 2003. Nucleotide diversity at two phytochrome loci along a latitudinal cline in *Pinus sylvestris*. *Molecular Ecology*. 12:1195–1206.

Gardner MF, Hechenleitner V P, Thomas P, Echeverria C, Escobar B, Brownless P, Martínez C. 2006. Threatened Plants of Central and South Chile: Distribution, Conservation and Propagation. Universidad Austral de Chile and Royal Botanic Garden Edinburgh. 188 pp.

Hair JB. 1966. Biosystematics of the New Zealand flora, 1945–1964. *New Zealand Journal of Botany*. 4:559–595.

Hair JB, Beuzenberg EJ. 1958. Chromosomal evolution in the Podocarpaceae. *Nature*. 181:1584.

Hamrick JL, Godt MJW, Sherman-Broyles SL. 1992. Factors influencing levels of genetic diversity in woody plant species. *New Forests*. 6:95–124.

Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*. 6.

Johnston HR, Keats BJB, Sherman SL. 2019. 12 - Population Genetics. In: Pyeritz RE, Korf BR, Grody WW, editors, Emery and Rimoin's Principles and Practice of Medical Genetics and Genomics (Seventh Edition), Academic Press, pp. 359–373. 7th ed edition.

Jombart T. 2008. Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*. 24:1403–1405.

Jombart T, Ahmed I. 2011. adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*. 27:3070–3071.

Jombart T, Devillard S, Balloux F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*. 11:94.

Kado T, Yoshimaru H, Tsumura Y, Tachida H. 2003. DNA variation in a conifer, *Cryptomeria japonica* (Cupressaceae sensu lato). *Genetics*. 164:1547–1559.

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I. 2015. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular ecology resources*. 15:1179–1191.

Lara A, Villalba R. 1993. A 3620-year temperature record from *Fitzroya cupressoides* tree rings in southern South America. *Science*. 260:1104–1106.

Marchelli P, Baier C, Mengel C, Ziegenhagen B, Gallo LA. 2010. Biogeographic history of the threatened species *Araucaria araucana* (Molina) K. Koch and implications for conservation: A case study with organelle DNA markers. *Conservation Genetics*. 11:951–963.

Marchelli P, Gallo L, Scholz F, Ziegenhagen B. 1998. Chloroplast DNA markers reveal a geographical divide across Argentinean southern beech *Nothofagus nervosa* (Phil.) Dim. et Mil. distribution area. *Theoretical and Applied Genetics*. 97:642–646.

Marchelli P, Gallo LA. 2004. The combined role of glaciation and hybridization in shaping the distribution of genetic variation in a Patagonian southern beech. *Journal of Biogeography*. 31:451–460.

Markgraf V, McGlone M, Hope G. 1995. Neogene paleoenvironmental and paleoclimatic change in southern temperate ecosystems - a southern perspective. *Trends in Ecology & Evolution*. 10:143–147.

Martinez Araneda C. 2011. Plant phylogeography in southern South America. Ph.D. thesis, University of Edinburgh.

Mosca E, Eckert AJ, Liechty JD, Wegrzyn JL, La Porta N, Vendramin GG, Neale DB. 2012. Contrasting patterns of nucleotide diversity for four conifers of Alpine European forests. *Evolutionary Applications*. 5:762–775.

Nahuelhual L, Carmona A, Lara A, Echeverría C, González ME. 2012. Land-cover change to forest plantations: Proximate causes and implications for the landscape in south-central Chile. *Landscape and Urban Planning*. 107:12–20.

Namroud MC, Beaulieu J, Juge N, Laroche J, Bousquet J. 2008. Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. *Molecular Ecology*. 17:3599–3613.

O'Connell LM, Mosseler A, Rajora OP. 2007. Extensive long-distance pollen dispersal in a fragmented landscape maintains genetic diversity in white spruce. *Journal of Heredity*. 98:640–645.

Paris JR, Stevens JR, Catchen JM, Johnston S. 2017. Lost in parameter space: a road map for stacks. *Methods in Ecology and Evolution*. 8:1360–1373.

Pastorino MJ, Gallo LA. 2002. Quaternary evolutionary history of *Austrocedrus chilensis*, a cypress native to the Andean-Patagonian forest. *Journal of Biogeography*. 29:1167–1178.

Pastorino MJ, Marchelli P, Milleron M, Soliani C, Gallo LA. 2009. The effect of different glaciation patterns over the current genetic structure of the southern beech *Nothofagus antarctica*. *Genetica*. 136:79–88.

Paula M, Leonardo G. 2006. Multiple ice-age refugia in a southern beech of South America as evidenced by chloroplast DNA markers. *Conservation Genetics*. 7:591–603.

Penner GA, Bush A, Wise R, Kim W, Domier L, Kasha K, Laroche A, Scoles G, Molnar SJ, Fedak G. 1993. Reproducibility of random amplified polymorphic DNA (RAPD) analysis among laboratories. *Genome Research*. 2:341–345.

Premoli AC, Kitzberger T, Veblen TT. 2000a. Conservation genetics of the endangered conifer *Fitzroya cupressoides* in Chile and Argentina. *Conservation Genetics*. 1:57–66.

Premoli AC, Kitzberger T, Veblen TT. 2000b. Isozyme variation and recent biogeographical history of the long-lived conifer *Fitzroya cupressoides*. *Journal of Biogeography*. 27:251–260.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics*. 155:945–959.

Prunier J, Verta JP, MacKay JJ. 2016. Conifer genomics and adaptation: at the crossroads of genetic diversity and genome function. *New Phytologist*. 209:44–62.

Quiroga MP, Pacheco S, Malizia LR, Premoli AC. 2012. Shrinking forests under warming: Evidence of *Podocarpus parlatorei* (pino del cerro) from the subtropical andes. *Journal of Heredity*. 103:682–691.

Quiroga PM. 2009. Contribución para la Conservación de Podocarpaceae del Sur de Sudamerica a Partir de Patrones Genéticos y Biogeográficos. Ph.D. thesis, Universidad nacional del Comahue Centro Regional Universitario Bariloche Laboratorio Ecotono.

Robles Ortiz C. 2003. Expansión y transformación de la agricultura en una economía exportadora: La transición al capitalismo en Chile (1850 - 1930). *Historia agraria: Revista de agricultura e historia rural*. pp. 45–80.

Ruiz E, González F, Torres-Díaz C, Fuentes G, Mardones M, Stuessy T, Samuel R, Becerra J, Silva M. 2007. Genetic diversity and differentiation within and among Chilean populations of *Araucaria araucana* (Araucariaceae) based on allozyme variability. *Taxon*. 56:1221–1228.

Salas C, Donoso PJ, Vargas R, Arriagada CA, Pedraza R, Soto DP. 2016. The Forest sector in Chile: An overview and current challenges. *Journal of Forestry*. 114:562–571.

Slatkin M. 1987. Gene flow and the geographic structure of natural populations. *Science*. 236:787–792.

Slatkin M. 1993. Isolation by distance in equilibrium and non-equilibrium populations. *Evolution*. 47:264.

Suharyanto, Shiraishi S. 2011. Nucleotide diversities and genetic relationship in the three Japanese pine species; *Pinus thunbergii*, *Pinus densiflora*, and *Pinus luchuensis*. *Diversity*. 3:121–135.

Torrejón F, Cisternas M, Alvial I, Torres L. 2011. Consecuencias de la tala maderera colonial en los bosques de alerce de Chiloé, sur de Chile (Siglos XVI-XIX). *Magallania*. 39:75–95.

Tóth EG, Tremblay F, Housset JM, Bergeron Y, Carcaillet C. 2019. Geographic isolation and climatic variability contribute to genetic differentiation in fragmented populations of the long-lived subalpine conifer *Pinus cembra* L. in the western Alps. *BMC Evolutionary Biology*. 19:1–17.

Veblen TT, Delmastro RJ, Schlatter JE. 1976. The conservation of *Fitzroya cupressoides* and its environment in southern Chile. *Environmental Conservation*. 3:291–301.

Vergara PM, Pérez-Hernández CG, Hahn IJ, Soto GE. 2013. Deforestation in central Chile causes a rapid decline in landscape connectivity for a forest specialist bird species. *Ecological Research*. 28:481–492.

Villagrán C. 1991. Historia de los bosques templados del Sur de Chile durante el Tardiglacial y Postglacial. *Revista Chilena de Historia Natural*. 64:447–460.

Waples RS, Gaggiotti O. 2006. What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*. 15:1419–1439.

Weir BS, Cockerham CC. 1984. Estimating F-Statistics for the analysis of population structure. *Evolution*. 38:1358.

Zonneveld BJ. 2012. Conifer genome sizes of 172 species, covering 64 of 67 genera, range from 8 to 72 picogram. *Nordic Journal of Botany*. 30:490–502.

Chapter 4

Natural regeneration of *Prumnopitys andina* (Poepp. ex Endl) de Laubenfels

4.1 Introduction

4.1.1 Natural regeneration

Seedling establishment and natural regeneration are essential in maintaining the long term ecological function and values of forest systems (Donoso and Nyland, 2005). This includes the replacement of individuals lost by mortality, as well as providing opportunities for adaptation and dispersal in light of environmental changes. Regeneration may occur naturally via seed, and also via vegetative reproduction, although the latter limits dispersal potential and capacity for adaptive responses.

Successful establishment of seedlings depends on many factors, including the intrinsic biological aspects of the species (ability to produce flowers, seed production, seed dispersal, growth-rate), the environmental conditions where the species occurs (light intensity, temperature, water supply, type of soil, competition), levels of environmental degradation (e.g. pressures such as harvesting, fire, land-use change) and major environmental change such as climatic change (including changes in temperature and water availability).

4.1.2 Factors that might affect natural regeneration in conifer species

A. Species biology

An important aspect influencing the dynamics of conifer populations is competition with angiosperms. Of particular importance here is the generally slow growth of conifers (at least during the juvenile phase) which can reduce their ability to compete (for space, light and nutrients) with faster-growing angiosperms (Shen and Nelson 2018, Coomes et al. 2005, Mallik 2003). This can lead to regeneration limitation of conifer species, in habitats with a high frequency of angiosperms.

B. Environmental degradation

Alteration of ecosystems (e.g. due to illegal logging, livestock grazing, or fire) also affects regeneration, usually retarding or stopping the establishment of new individuals within the forest (Oldén et al. 2017, Borg et al. 1988), which can result in a significant limit to regeneration. However, for some tree species disturbance can have a positive effect and lead to increased regeneration. For instance, fires can control seedbed quality and enhance regeneration, and forestry practices which lead to the creation of canopy gaps or open land might also promote regeneration (Carswell et al. 2007, Beveridge 1973, Smith-Ramírez 2007). There is thus often an optimal level of disturbance for the establishment of new individuals within the forest.

C. Climatic variability

Variation in climate (temperature and precipitation) can also affect the replacement and establishment of new individuals (seedlings-saplings) (Madsen, 1995). This because the early developmental stages of trees (including the germination process) can be more susceptible to climate variability than adult stages (Walck et al., 2011) (e.g. some seeds might need long-cold exposure to germinate, seedlings might need humid conditions to survive). In this context, some investigations have given particular attention to correlating the latitudinal distribution of some species and their response to seedlings establishment (e.g. Gao et al. 2017, Wei et al. 2015), as seedlings might be significantly influenced by climate variability (temperature and water supply) along such gradients (Bognounou et al., 2010). For instance, Matías and Jump (2015) assessed demographic structure and regeneration of two conifer species; *Pinus sylvestris* L. and *Juniperus communis* L. The authors observed a more limited regeneration of the two species at the southern or northern limit of their distribution, linking this pattern with the increase of temperature in these areas (temperature increases compared to past at the range margin).

4.1.3 Introducing *Prumnopitys* as a case study

Prumnopitys andina (Poepp. ex Endl) de Laubenfels is an endemic conifer, restricted to central-southern Chile, and represent the southernmost species of the genus in South America (5 species in total). The existing area of occupancy of *Pr. andina* is estimated to be about 36 km^2 (Gardner, 2013), with only 12 locations identified and some of them with <50 adult individuals. The current patchy distribution of *Prumnopitys* is mostly attributed to recent disturbances, including dams changing watercourses, increased fire frequency, landuse-change and illegal logging. A significant increase in temperatures and a decrease in rainfall have also been identified as threats that the species has suffered over the last decades (mostly in the northern central part of its distribution) (Williams, 2017). Therefore *Pr. andina* represents an important conservation case study to understand the impacts of modifications to its natural ecosystem on regeneration. Yet despite its ecological importance, and its conservation value, there have been no investigations on patterns of regeneration in this species.

4.1.4 Objectives

The focus of this chapter is to assess levels of regeneration of *Pr. andina* throughout its entire natural distribution. A secondary aim is more general documentation and characterisation of the attributes of populations across its distribution, including the main threats observed in each of the sites visited. More specifically, I aimed to investigate factors that might affect natural regeneration of this species, by measuring seedling and sapling density and exploring its association with 1) seedling competition, 2) *Pr. andina* adult population density 3) light levels, and 4) the latitudinal distribution of the species. This statistical assessment of seedling and sapling density is then complemented with field observations of the population size and characteristics, and documentation of sources of threats to the species.

Due to time constraints, it was not possible to extend this work to the other three species studied in this Thesis. Nevertheless, I also report (in the Discussion 4.4) informal observations about levels of regeneration in *Saxegothaea conspicua* Lindley, *Podocarpus salignus* D. Don and *Fitzroya cupressoides* (Molina) Johnston. This to provide addi-

tional context on regeneration in other conifer species from the temperate rain forest of Chile.

4.2 Methods

4.2.1 General information on *Pr. andina*

Prumnopitys andina is a dioecious (rarely monoecious) endemic conifer species from Chile. It shows a very narrow distribution in the central-southern part of the country from 35°52' to 39°08' S with a range of altitude between 300 - 1000 m, most commonly occurring between 700 - 900 m. There are approximately 12 known locations (which will also be referred to as populations in this chapter). Most of these populations are in the eastern part of the country, in the Andean mountain range, where they typically occur on the slopes of valleys near watercourses. A single population is located in the Coastal range, in the Araucanía region (Nahuelbuata, 37°48' S) and another in the Central valley (Victoria village, Araucania region, Pua-Santa Lucia, 38°20' S).

Prumnopitys andina is a shade-tolerant species with a height up to 20 m, and about 0.5 to 1 m diameter (Donoso Zegers, 2006). It occurs in humid areas and soils rich in nutrients. The species is commonly associated with *Nothofagus obliqua* Mirb., *Nothofagus dombeyi* Mirb. Oerst., *Nothofagus alpina* Popp. & Endl. and also with other species such as; *Austrocedrus chilensis* (D.Don) Pic.Serm. & Bizzarri, *Quillaja saponaria* Molina, *Lomatia hirsuta* Diels ex J.F.Macbr, among others (Donoso Zegers, 2006)

Prumnopitys andina produces an edible green fruit which falls by gravity. Each fruit has a single seed inside. Livestock and rodents are commonly attracted to the fruits, and they might act to disperse the seeds over short distances. Observations at the RBGE have shown that *Pr. andina* has extremely variable seed germination rates (unpublished work), despite informal personal field observations suggest high levels of seed viability. The species seems to be very successful in regenerating vegetatively in greenhouse conditions. However, the extent of clonality in the wild is unknown (Donoso Zegers, 2006).

4.2.2 Study area

This investigation covered a total of ten populations, representing the entire natural range distribution of *Pr. andina* (Figure 4.1). Eight of the ten sites included in this study are in the Andes range with distribution from the Maule until the Araucanía Region (35-39° S). The most northern population studied is Corral de Salas (the northern limit of the species), and likewise, the most southerly population at Reigolil was also included. The population with the lowest elevation is the Central valley population in Pua-Santa Lucia at 395 m. The only existing coastal population (Nahuelbuta) was also visited. However, due to the small size of this population, with trees extremely scattered along the valley and the challenging topography of the area, no empirical data were recorded. Instead, I took informal notes from this population.

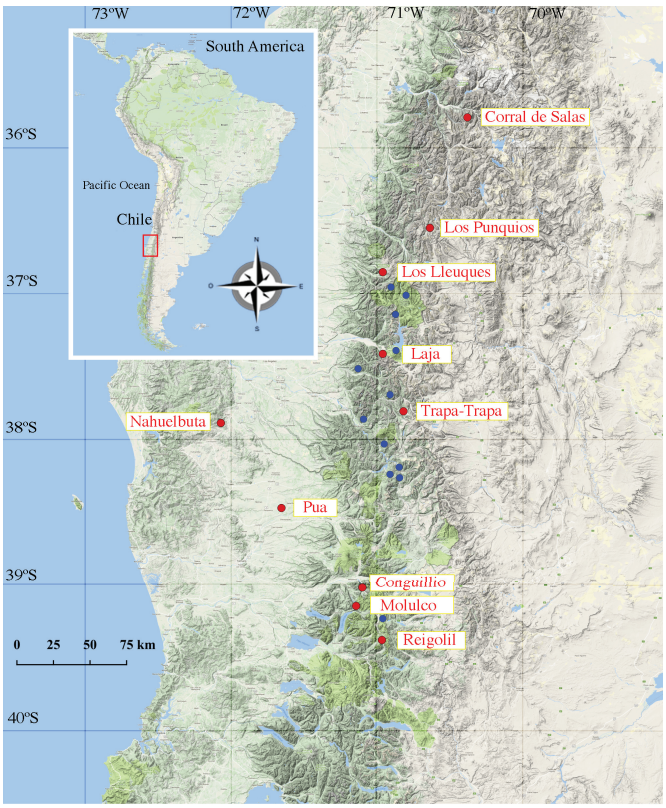


Figure 4.1: Map showing the entire natural distribution of *Pr. andina* (blue dots) and the populations included in this investigation (red dots).

Table 4.1: Populations included in the study, including their region of origin, location, elevation, area and an estimation of the number of adult individuals of *Pr. andina*. The populations are listed from the north to the south. (Area of each plot was estimated using google earth).

nº plots	Stand name	Region	Latitude (S)	Longitude (W)	Altitude (m)	Area (ha)	nº adults trees
3	Corral de Salas	Andes	35°52'58"	70°59'22"	1017	6	40-50
10	Los Punquios	Andes	36°32'15"	71°11'47"	963	104	600-800
9	Los Lleuques	Andes	36°51'40"	71°36'16"	957	8	150-200
10	Laja	Andes	37°22'25"	71°33'53"	833	6	50-70
10	Trapa-Trapa	Andes	37°40'13"	72°01'12"	788	101	500-700
obs	Nahuelbuta	Coastal	37°49'30"	72°48'29"	871	-	200*
10	Pua-Santa Lucia	Central valley	38°20'26"	72°19'09"	395	4	100-200
10	Conguillío	Andes	38°49'36"	71°39'44"	579-1002	16	500-600
10	Molulco	Andes	38°56'9"	71°42'36"	744	31	400-500
10	Reigolil	Andes	39°08'08"	71°28'58"	826	301	>3000

*The number of *Pr. andina* individuals indicated in the Nahuelbuta Coastal population is based on the census done by the forestry company (Arauco) which are the land-owners where the population is located. However my personal observations suggest a much smaller population (less than a 100 trees).

4.2.3 Natural regeneration of the *Pr. andina* forest

To measure natural regeneration, eight populations were sampled using ten plots of 10x10 m (100 m^2), which is also referred to as a 'stand' in this chapter. This sampling strategy was not possible in two of the populations: the northern population (Corral de Salas) was surveyed only with three plots of 10x10 m due to small population size and an extremely scattered presence of *Pr. andina*. The Los Lleuques population was studied with nine plots (4.1). Each plot was gridded and marked every two meters with stakes. Furthermore, ten subplots of 1x1 m (1 m^2) were systematically established inside each main plot. Each subplot was gridded with a thin wood frame (1 m x 1 m), and the survey was undertaken inside each subplot (Figure 4.2), quantifying the number of seedlings and saplings of *Pr. andina*, and the presence of other species. The criteria for scoring an individual as a seedling was <2 cm DBH and <1 m tall. For saplings it was <2 cm DBH and >1 m tall. The species identifications were undertaken *in situ*. Seedlings and saplings that were not possible to identify were categorised as 'sp'.

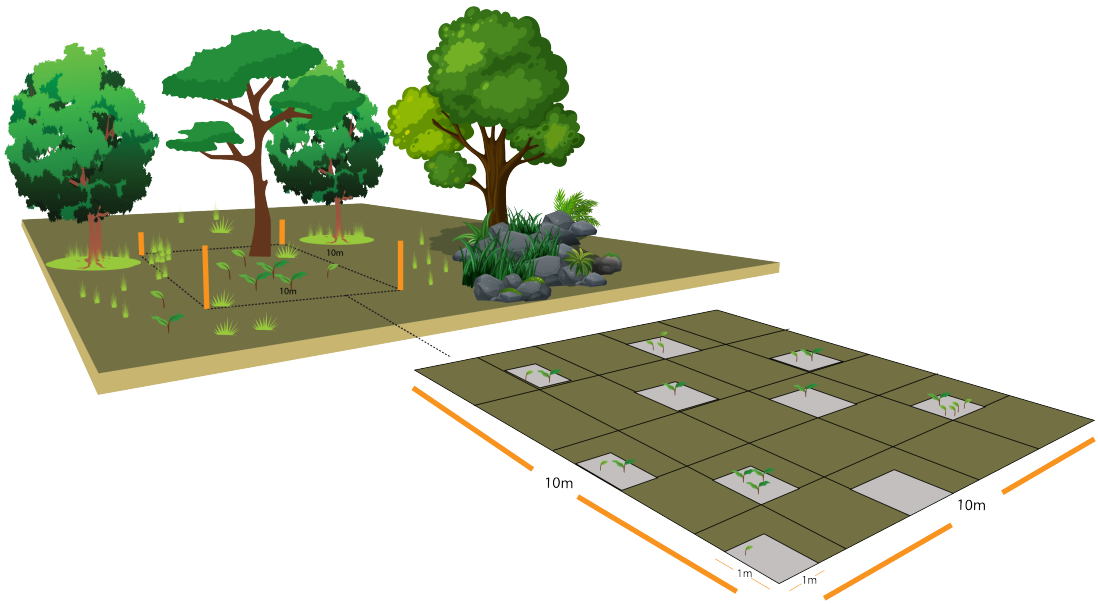


Figure 4.2: Graphic illustration of the plots and subplots to quantify seedling and sapling density of the *Pr. andina* forest. The top image is a representation of the area and plot where the survey was done. The image below represents a projection of a standard plot of 10x10 m made to evaluate seedling and sapling density of *Pr. andina*. The ten grey small square of 1x1 m in the image below, represent the subplots where the seedling and sapling density quantification was done.

4.2.4 Factors potentially associated with the natural regeneration of the conifer species

To explore the factors associated with the natural regeneration, I tested whether there was a correlation between the number of seedlings and saplings of *Pr. andina* and the number of seedlings and saplings of other species (e.g. to evaluate seedling competition). I also tested if the number of seedlings and saplings of *Pr. andina* was associated with the number of *Pr. andina* adults, the level of sunlight present in each stand and if there was a correlation between number of seedlings and saplings and the latitudinal distribution of the species.

To undertake these tests, I firstly conducted a test of normality for all variables, including a matrix scatter plot to observe the distribution of the data, in order to choose the appropriate correlation test (Pearson or Spearman). This was followed by a bivariate correlation estimated using the SPSS statistics software version 23.

- The total number of *Pr. andina* adults present in each of the stand-population was quantified by sex (female-male-monoecious).
- Sunlight was measured in each stand-population following the methodology of Parker and Donoso (1993). This methodology is based on observations of light intensity through the forest canopy.

4.2.5 Characterisation of the *Pr. andina* forest

To complement the information recorded from each stand-population and to characterise the *Pr. andina* forest, the following data were also recorded; the location, an estimation of the population size (ha) and an estimation of the total number of mature trees in each of the populations (only *Pr. andina*). The evaluation of the population size was made recording the boundaries of each site with GPS to then estimate the population size using Google Earth (Table 4.1). The estimation of the number of mature trees of the conifer species was made visually by a team of four people. The main threats were also recorded following the methodology of threat assessment provided by the IUCN (see Appendix D for further information).

A species list was also recorded in each of the stand-populations using Braun- Blanquet's Cover Abundance Scale: + (trace); 1 (1-5%); 2 (2-25%); 3 (25-50%); 4 (50-75%); 5 (75-100%). The species list focused on trees, shrubs, herbs and climbers plants, excluding ferns due to their low occurrence in the Chilean coniferous forests. For better interpretation about the cover-abundance of each species, Braun-Blanquets Cover Abundance values were converted to Midpoint coverage range (see conversion in Appendix, Table C.5). The size distribution of the largest trees in each stand-population was also recorded (DBH-height) (for *Pr. andina* or for other species if the largest trees were of a different species). This gives a picture of the size of the largest trees based on a sample size of 2 trees x 10 stands (n= 20) in most populations except for two where only 3 (total n= 6) and 9 (total n= 18) stands were assessed (see Table 4.1). This recording of the two largest trees (of any species) is designed as a general measure of forest maturity/size.

4.3 Results

I present the results of this study by firstly describing the *Pr. andina* populations, and their characteristics, I secondly report levels of seedling and sapling density over the different sites, I thirdly assess the relationship between levels of seedling and sapling density and the characteristics of the sites and finally provide more detailed observations at the population level.

4.3.1 Characteristics of the *Pr. andina* forest

Presence of *Pr. andina* adults in each population

The largest number of *Pr. andina* trees (adults) across the stand-populations were found in Conguillio (in the south of the species range) with a total of 98 individuals (total area surveyed=1000 m^2). This population had the largest tree density across the stands, with a total of 0.098 trees per m^2 (Table 4.2). The second-largest tree density of *Pr. andina* was also found in the southern populations (Molulco and Reigolil with 0.069 and 0.60 trees/m^2 respectively). The lowest number of *Pr. andina* adult was found in the northern population Corral de Salas (0.017 trees/m^2) followed by Laja (0.019 trees/m^2), Los Lleuques (0.02 trees/m^2) and Los Punquios (0.022 trees/m^2), all populations from the central-northern part of the country. Overall, most of the populations showed a high proportion of male individuals of *Pr. andina*, except in the southern population Reigolil, which showed a higher density of females trees than males. However, it was not possible to identify the sex of a high number of individuals in some populations, due to the lack of the cones.

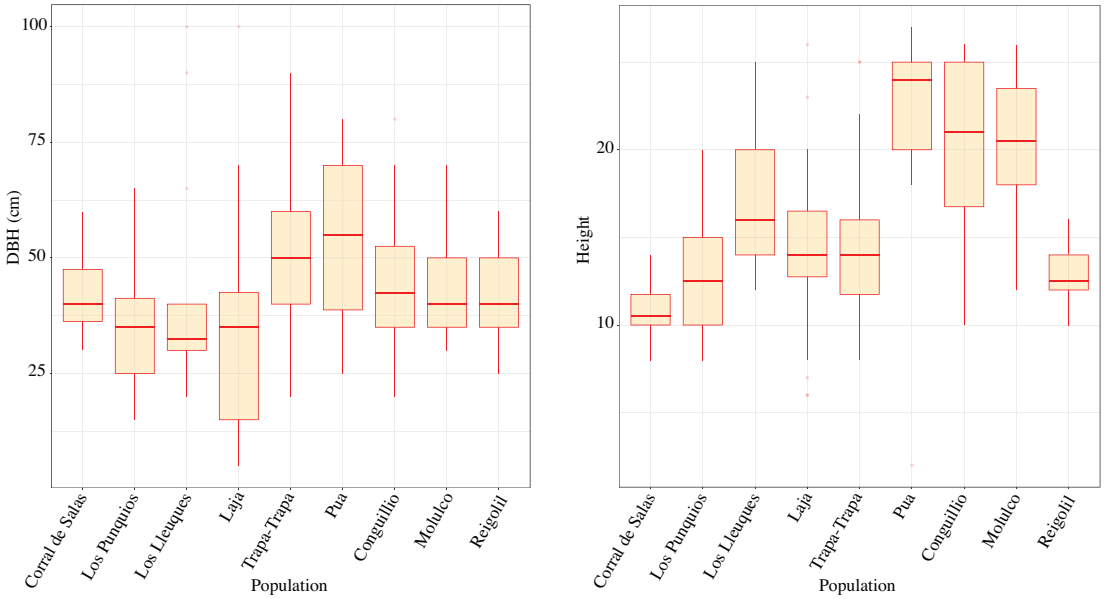
Table 4.2: Number and density of *Pr. andina* adults found by sex across the stand-populations. Populations are listed from the north (top) to the south (bottom). *Two monoecious individuals were identified in the Reigolil population and added to the total number of adults (Total adults).

Population	nº plots	Total male	Total female	Total unknown	Total adults	size plots (m^2)	Total density	density male	density female	density uknown
Corral de Salas	3	5	0	0	5	300	0.017	0.017	0	0
Los Punquios	10	11	6	5	22	1000	0.022	0.011	0.006	0.005
Los Lleuques	9	9	5	4	18	900	0.020	0.010	0.006	0.004
Laja	10	7	6	6	19	1000	0.019	0.007	0.006	0.006
Trapa Trapa	10	25	15	8	48	1000	0.048	0.025	0.015	0.008
Pua	10	5	1	38	44	1000	0.044	0.005	0.001	0.038
Conguillio	10	4	3	91	98	1000	0.098	0.004	0.003	0.091
Molulco	10	0	7	62	69	1000	0.069	0	0.007	0.062
Reigolil*	10	5	23	30	60	1000	0.060	0.005	0.023	0.030

Size distribution of the largest trees in each population

Overall, the size distribution (DBH and height) of the largest trees (based on the two largest individuals found within each of the stand-populations), showed a median DBH=40 cm and height=16 m (Figure 4.3, full list in Appendix, Figure C.1). The trees with the largest DBH were found in the central part of Chile in Pua-Santa Lucia (Central valley population) and Trapa-Trapa. The tallest trees were found in the southern populations (Pua-Conguillio- Molulco).

The populations where *Pr. andina* was recorded as the most prominent trees (within the two tallest trees recorded in each stand-population), were in Corral de Salas (northern population), Laja, Trapa-Trapa (central south population), Santa Lucia-Pua (central valley population) and in Reigolil (southern population, see Illustration in Figure 4.4).



(a) DBH (b) Height

Figure 4.3: DBH and height distribution of the largest individuals across stand-populations. Populations listed from the north (left) to the south (right). Full list in Appendix, Figure C.1

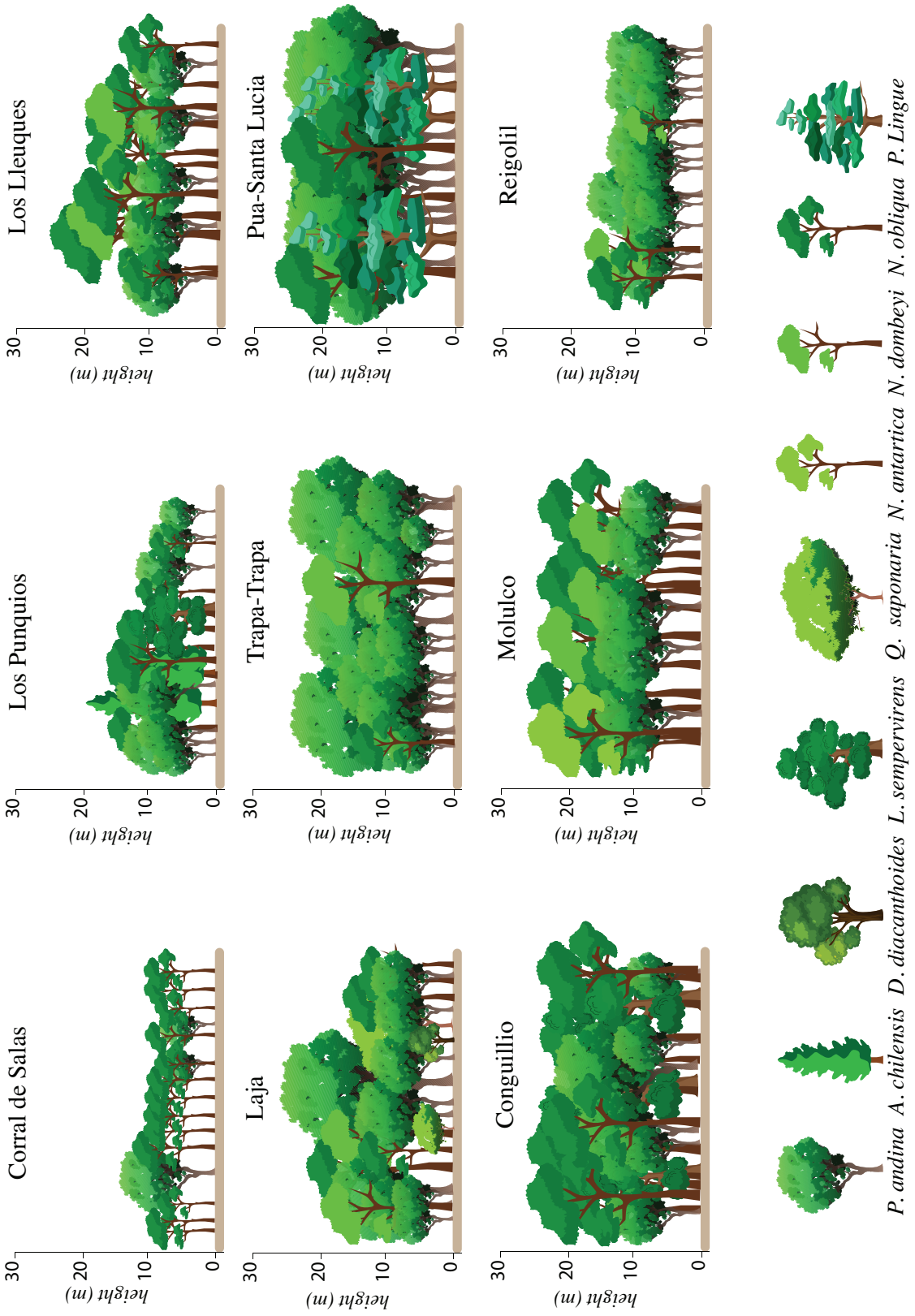


Figure 4.4: Illustration of the forest structure of each population of *Pr. andina*. This illustration is based on the tallest trees recorded in each population. Populations are listed from the north (top-left) to the South (bottom-right).

Associated species found in each population

The vegetation associated with *Pr. andina* was diverse. A total of 42 species in 25 families were observed across populations (excluding non- vascular plants, full list in Appendix, Table C.2). The highest proportion of species by life form corresponds to trees (57%) followed by shrubs (32%) and finally herbs and climbers species with less than 5% (Figure 4.5). Across all populations, *Pr. andina* had the highest cover-abundance, followed by *N. obliqua* and to a lesser extent *L. hirsuta* (Figure 4.6). However, an exception was found in Corral de Salas and Laja populations where the conifer was not the most abundant species. (Appendix, Table C.4).

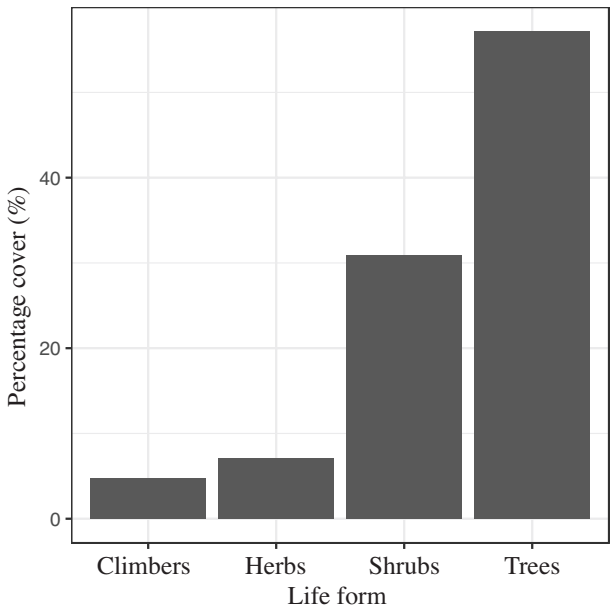


Figure 4.5: Percent of species by life form associated with the *Pr. andina* forest.

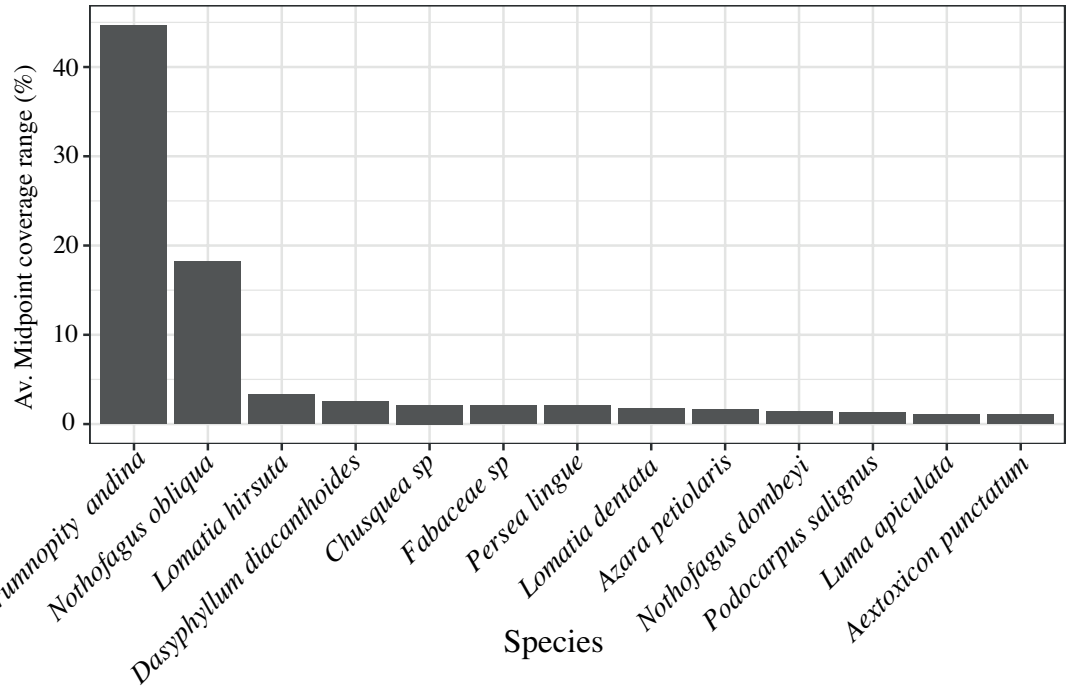
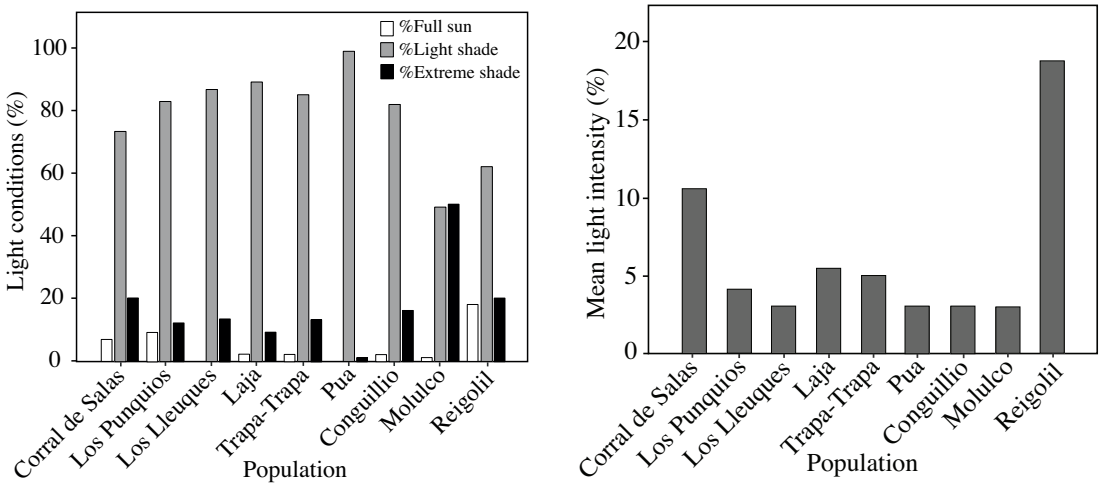


Figure 4.6: The ten most abundant species in *Pr. andina* forests across all populations, using Braun-Blanquet scale to the Midpoint of the coverage range. Full list in Appendix, Table C.3.

Light intensity in each population

Overall, populations were characterised by a canopy with light-shade conditions (dappled, trees provide high proportion of canopy closure and plants do not receive direct sun) (Parker and Donoso, 1993). Indeed, most of the stands surveyed in each population showed light-shade as the pre-dominant light condition and to a lesser extent, extreme shade and full sun conditions (Figure 4.7). The highest mean light intensity was found in Reigolil with about 18% (the southern population) followed by Corral de Salas with about 10% (the northern population). The other populations observed a mean light intensity between 4-6% (Figure 4.7).



(a) (b)

Figure 4.7: Light intensity by population. (a) Proportion of light conditions by population. (b) Mean light intensity estimation by population following Wellner (1979)

4.3.2 Levels of natural regeneration in the *Pr. andina* forest

The trend observed in most populations was characterised by seedlings (99%) rather than saplings. The greatest seedling density observed was in Los Lleuques, with 987 seedlings recorded equivalent to 0.99 seedlings per m^2, of which 74% were seedlings of *Pr. andina* and 26% seedlings from other species. The Conguillio population registered the second greatest level of seedling density, with 837 seedlings equivalent to 0.84 seedlings per m^2. However, only 31% of seedlings were of *Pr. andina*. The other greatest seedling density was found in Reigolil (599 seedlings) which represent the southern and the largest population throughout the entire natural distribution of the species. In this population, 82% of the seedling density were of *Pr. andina*. The lowest level of seedling density was in the northern populations in Corral de Salas (with no evidence of *Pr. andina* regeneration), followed by Trapa-Trapa where most of the seedling density was associated with other species (Table 4.3).

Table 4.3: Total seedling and sapling of *Pr. andina* by population. Full distribution of seedlings (*Pr. andina* and other species combined) in Appendix, Figure C.1. Density was calculated only with seedlings.

Population	Total covered (m2)	*Pr. andina* seedling (nº)	*Pr. andina* sapling (nº)	*Other sp* seedling (nº)	*Other sp* sapling (nº)	Total seedling density: Total nº	Total seedling density: All sp combined (m2)	Total seedling density: *Pr. andina* (m2)	% seedling density: *Pr. andina*	% seedling density: *Other sp*
Corral de Salas	300	0	0	27	5	27	0.090	0	0	100
Los Punquios	900	89	0	128	7	217	0.241	0.098	40	59.9
Los Lleuques	1000	732	0	255	6	987	0.987	0.732	74.2	25.8
Laja	1000	40	0	327	7	367	0.367	0.040	10.9	89.1
Trapa-Trapa	1000	9	0	136	0	145	0.145	0.009	6.2	93.8
Pua	1000	49	0	501	1	550	0.550	0.049	8.8	91.2
Conguillio	1000	259	17	578	4	837	0.837	0.259	30.9	69.1
Molulco	1000	154	0	100	0	254	0.254	0.154	60.6	39.4
Reigolil	1000	493	7	106	0	599	0.599	0.493	82.3	17.7

4.3.3 Statistical association between levels of regeneration and the stand characteristics

The level of competition (seedling density associated with other species), the presence of adults of *Pr. andina* in each population, the level of luminosity and the latitudinal gradient, did not show a normal distribution (details in Appendix Table C.7). Thus, due to the non-normal data distribution, the non-parametric Spearman Rank Order Correlation Coefficient was used to test for associations of the different variables with levels of regeneration. A statistical association with saplings were not included due to its low occurrence.

This Spearman Rank Order Correlation approach showed a significant positive association between the number of seedling with the total number of *Pr. andina* adults present in each population (Coefficient= 0.329, p= 0.003). Similarly, a significant but negative association was found when the number of seddlings was correlated with the number of male adults (Correlation Coefficient= -0.229, p= 0.038, p= <0.05, Table 4.4). In contrast, there was no significant association between the number of seedling when correlated only with the number of female adults of the species. However, the p-value was borderline significant (Correlation Coefficient= 0.209, p= 0.059). A significant positive correlation was also found between the number of seedling and the latitudinal distribution of the species (Coefficient= 0.438, p= 0.000). However, there were no significant associations found with the abundance of seedlings of other species

or with the light intensity observed across populations (Table 4.4).

Table 4.4: Spearman's test correlation coefficient for each factor potentially associated with variation in the number of seedlings. Female, Male and Total Adults refer only to *Pr. andina*. Total seedling refers to the number of seedlings of *Pr. andina* and other species. *Correlation is significant at the 0.05 level (2-tailed).**Correlation is significant at the 0.01 level (2-tailed).

		***Pr. andina* seedling**	**Female**	**Male**	**Total Adults**	**Light intensity**	**Seedling other species**	**Total seedling**	**Latitudinal gradient**
***Pr. andina* seedling**	Correlation Coefficient	1	0.209	-0.229*	0.329**	-0.144	0.170	0.703**	0.438**
	Sig. (2-tailed)	.	0.059	0.038	0.003	0.197	0.126	0	0.000
	N	82	82	82	82	82	82	82	82
Female	Correlation Coefficient	0.209	1	0.207	0.175	-0.075	-0.097	0.155	0.076
	Sig. (2-tailed)	0.059	.	0.062	0.115	0.501	0.388	0.166	0.499
	N	82	82	82	82	82	82	82	82
Male	Correlation Coefficient	-0.229*	0.207	1	0.046	-0.026	-0.034	-0.078	-0.237
	Sig. (2-tailed)	0.038	0.062	.	0.681	0.816	0.763	0.487	0.032
	N	82	82	82	82	82	82	82	82
Total Adults	Correlation Coefficient	0.329**	0.175	0.046	1	0.046	-0.142	0.108	0.572
	Sig. (2-tailed)	0.003	0.115	0.681	.	0.682	0.202	0.336	0.000
	N	82	82	82	82	82	82	82	82
Light intensity	Correlation Coefficient	-0.144	-0.075	-0.026	0.046	1	-0.336**	-0.357**	0.330
	Sig. (2-tailed)	0.197	0.501	0.816	0.682	.	0.002	0.001	0.765
	N	82	82	82	82	82	82	82	82
Seedling other species	Correlation Coefficient	0.170	-0.097	-0.034	-0.142	-0.336**	1	0.733**	-0.105
	Sig. (2-tailed)	0.126	0.388	0.763	0.202	0.002	.	0	0.347
	N	82	82	82	82	82	82	82	82
Total seedling	Correlation Coefficient	0.703**	0.155	-0.078	0.108	-0.357**	0.733**	1	0.172
	Sig. (2-tailed)	0	0.166	0.487	0.336	0.001	0	.	0.122
	N	82	82	82	82	82	82	82	82
Gradient	Correlation Coefficient	0.438**	0.076	-0.237	0.572	0.330	-0.105	0.172	1
	Sig. (2-tailed)	0.000	0.499	0.032	0.000	0.765	0.347	0.122	.
	N	82	82	82	82	82	82	82	82

4.3.4 Summary of the main levels of threat

At each of the ten locations of *Pr. andina* visited the significant threats were recorded. The major factors impacting (or potentially impacting on populations) were as follows:

1. **Agriculture which includes, wood and pulp plantations and livestock farming**.

    - Wood and pulp plantations were observed associated with 30% of the populations. They represent a threat, in the sense that (a) They can reduce population sizes, and it seems likely in many cases that populations of *Pr. andina* were impacted by the planting of exotic species (b) They can introduce competition. Commercial plantations are usually of rapidly-growing species (*Pinus* and *Eucalyptus*) which can have substantial impacts on the natural ecosystems of *Pr. andina*, including; the reduction of water supply, soil acidification and soil impoverishment.
    - Livestock farming were observed in all sites surveyed. This reflects a threat mainly because exotic animals (livestock, pigs, sheep) are usually free within the forests, trampling and browsing seedlings and saplings of any native species. Both threats were considered as a medium and high-impact.

2. **Biological resource use including logging and wood harvesting**.

    - Logging and wood harvesting were observed in about 40% of the populations surveyed. This represents a threat, in the sense that (a) Illegal logging on *Pr. andina* degrade the forest, reducing the number of reproductive adults, (b) Illegal logging of associated species (cause damaged on seedlings and saplings during wood extraction. This threat was categorised a low to medium-impact.

3. **Residential and commercial development (human disturbance) which includes, housing and urban areas, and tourism and recreation activities**.

    - The presence of housing and urban areas was observed in around 40% of the populations, and tourism and recreation activities were perceived at 30% of the sites surveyed. They represent a threat, in the sense that (a) trees are felled to create pathways along the riverside, (b) trees are cut for firewood for campfires, (c) fires spreading from campfires can destroy trees, and (d) disturbance from humans and vehicles results in seedlings being trampled, restricting regeneration. However, overall, the scale of threats from residential and commercial development is low, as in all cases, the impacts observed were limited to small portions of the population, and even in the impacted areas, the consequences appear to be minor.

4.3.5 Notes on individual populations

Northern populations

Corral de Salas. This is the northernmost population of *Pr. andina* located at 35° S. It is a tiny population, 5.44 ha in area, spread over about 1.5 km along the valley and no more than 50 adult individuals of the species. The forest is dominated by *N. obliqua* with *Pr. andina* disparately distributed in the dense *N. obliqua* forest along the main river that goes through the valley. There is evidence for the presence of threat from grazing due to the presence of cattle that usually browse herbaceous, seedlings, buds and pastures in the area. No other clear threats were observed (e.g. logging, exotic plantations, hydroelectric scheme).

Central populations of the Andes

The central populations which include; Los Punquios, Los Lleuques, Laja and Trapa-Trapa (listed from the north to the south) are all distributed on the slopes of the Andes range between 36-37° S.

The Punquios. The population is characterised by a mixed forest on the eastern side of the Sauces and Ñuble river, generally composed of stony soil with some places with a surface of leaf mould. The population area is ca 103 ha, over a distance of about 26 km and with about 800 adult individuals of the species. *N. obliqua* is the dominant species, with other species present including *A. chilensis*, *Q. saponaria*, *Laurelia sempervirens* (Ruiz & Pav.) Tul., among others. *Pr. andina* is scattered along the valley but is more abundant close to watercourses or ravines. In these locations *Pr. andina* forms pure forest with low competition and free understory. This population is highly threatened by animals, mainly by cattle, sheep and pigs. Evidence of logging is present in the area; however, these activities are primarily targeted at Nothofagaceae species

The Lleuques. This tiny fragmented population is divided into three areas. The entire population is about 7.59 ha, spread over about 1.63 km and containing about 200 adult individuals of the species. One of the areas is associated with a dense forest of *N. obliqua* and other species such as; *Nothofagus antarctica* (Forster) Oerst, *Aextoxicon punctatum* Ruiz & Pav, *Maytenus boaria* Molina, *N. alpina*, *P. salignus*, *Embothrium coccinea* J.R. Forst. & G. Forst and *L. hirsuta*. In this area, *Pr. andina* is scattered through the mixed forest. The other two areas of less than 1 ha each are characterised by small patches of pure *Pr. andina* forest, on flat areas with slopes of less than 5%. The soil is composed mainly of leaf mould. There was evidence of grazing animals throughout the entire population, and also evidence of extensive logging, mostly of *N. obliqua*. Howver, there was some evidence of logging of the *Pr. andina* as well.

Laja: Tiny population, 5.9 ha, spread over about 1.1 km with no more than 70 adult individuals. *Pr. andin*a is very dispersed along the main river. The area is dominated by a mixture of species including *A. punctatum*, *Dasyphyllum diacanthoides* (Less.) Cabrera, *Luma apiculata* (DC.) Burret, *L. sempervirens*, *Gevuina avellana* (Molina) Gaertn., *N. dombeyi*, *N. antartica*, *M. boaria*, *A. chilensis* and *Fuchsia magellanica* Lam. The main threats noted were disturbances associated with cattle and in the past of land-use change, as the population is surrounded by a large exotic conifer population which could have reduced the population size of *Pr. andina*.

Trapa-Trapa: This population covers an area of about 103 ha and a distance of about 14 km long. The population has about 700 adult trees. However, the individuals are distributed in patches confined to the main valleys. Large old-growth trees are typically distributed near watercourses in areas with slopes more than 5%. The area shows sandy soil and some rocky areas as well. The vegetation was sparse, and associated species include *N. obliqua*, *Fabiana imbricata* Ruiz & Pav, and *Berberis sp.* The main threats observed related to disturbances from sheep and pigs in the area. There is also evidence of logging of the native forest (including few individuals of *Pr. andina*)

Coastal populations

Nahuelbuata. The only existing coastal population of *Pr. andina* is located at 37° S, outside the National Park Nahuelbuata. This is a tiny population with a few individuals scattered among exotic plantations (*Pinus radiata* D.Don and *Pinus oregon* (Mirbel) Franco and *Eucalyptus globulus* Labill). This population is under a high risk of disappearing. According to the forestry company (Arauco, land-owners), the number of individuals is about 200. However, after extensive observations, my survey suggested the population was smaller, with no more than 100 trees. Apart from the exotic plantations, other associated disturbances in the area are animals (mainly cattle). The soil is dry and seemed poor in nutrients

Central valley populations in the south of Chile

Pua-Santa Lucia. The only extant Central valley population of *Pr. andina* is located at 38° S at Pua-Santa Lucia. This is a tiny and fragmented population of 4 ha, spread over about 1.36 km, characterized by two small patches of about 2 ha each (1km distance between them) with around 100-200 large old-growth trees in each patch. Although *Pr. andina* is locally abundant it is associated with other co-dominant species such as *Persea lingue* (Ruiz & Pav.) Nees ex Kopp, *N. obliqua* and *L. sempervirens*. The *Pr andina* plants occur on soil dominated by leaf mould, on generally flat ground. The two fragmented population are surrounded by a huge agriculture area growing pasture for cattle and fruit trees. There is a massive presence of livestock and also evidence of logging in the area. This suggests that the population may once have been larger, and certainly that any future expansion would be constrained by intensive surrounding land use.

Southern populations

The southern populations include Conguillio, Molulco and Reigolil (listed from the north to the south) and are located between 38-39° S, in the foothills of the Andean mountains. Weather conditions here are wetter and with shorter summer (December-March) than the northern populations (Donoso Zegers, 2006). The Conguillio population is the only one located in a national park in that area. Reigolil is the southernmost population of *Pr. andina*.

Conguillio: This medium-size population, of 16.2 ha, is about 15 km long with around 500-600 large old-growth *Pr. andina* trees. The population is divided into two or three subpopulations, one of which is lowland with *Pr. andina* growing along the main road to the National Park Conguillio, on a flat terrace plain with some areas with no more than 5% slope with leaf mould soil. The vegetation is composed of *L. sempervirens*, *Laureliopsis philippiana* (Looser) R. Schodde, *N. obliqua*, *G. avellana*, *L. dentata*, *Sophora microphylla* Aiton, *L. hirsuta*, *Aristotelia chilensis* (Molina) Stuntz, *Rhaphithamnus spinosus* Miers and *Azara sp.* A second subpopulation is located just

in the entrance of the National Park (around 6 km from the first one) where *Pr. andina* is more abundant. The vegetation and soil composition are similar to the first area described. The last subpopulation is located next to Laguna Verde (around 9 km from the second area) inside of the National Park Conguillio. The area is characterised by old forest with some places dominated by *Pr. andina* and other dominated by *N. obliqua*, and there is a dense canopy throughout with the plants growing on volcanic soil. The number of *Pr. andina* individuals in this subpopulation are considerably less than the first two areas. No specific threats were observed in this sub-population, inside the national park. However, the individuals that are outside the national park show evidence of impacts from cattle and logging. Volcanic activities are also common in this area, and there is a risk that volcanic activity might result in impacts on this population.

Molulco: The population here is about 35 ha in area and spread over 1 km with around 400-500 adult trees. The forest is mostly dominated by *N. obliqua* and *N. dombeyi*. However, some areas are dominated entirely by *Pr. andina*, forming a pure forest of the species. The soil is normally composed of leaf mould, and the slope gradient is not higher than 10%. Field observations indicate wood-cutting, mostly of *N. obliqua* and *N. dombeyi*. Livestock is frequently present, and grazing may present a threat to *Pr. andina* regeneration.

Reigolil: This is a massive population, covering 301 ha with more than 3000 adults' individuals of the species. The population extends from the Reigolil village southwards for about 30 km along the river. Individuals of *P. andina* are mostly located next to the main river or small watercourses along the valley. The main associated flora consists of *N. obliqua* and *N. dombeyi*, *A. chilensis* and in the southern area *S. conspicua*. *Prumnopitys andina* form pure forest in some areas of this population. The population shows compact soil with some rocky areas. The main disturbances are associated with cattle, pigs, sheep, and there is also some evidence of logging.

4.4 Discussion

Natural regeneration of the *Pr. andina* forest

In this investigation, we found that seedling of *Pr. andina* was usually frequent throughout the natural distribution of the species, except for the northernmost (Corral de Salas), and the only coastal population (Nahuelbuta). Additionally, the number of seedlings of *Prumnopitys* was strongly associated with the number of adults trees and latitudinal gradient. In contrast, there was no statistical association with the level of seedlings competition (from other species) and light intensity. However, light shade condition was frequent in all stands where regeneration was observed.

In the first instance, the number of seedlings of the conifer was associated with the presence of adults trees (*Pr. andina* trees) within a population, as *Prumnopitys* generally disperse their seeds by gravity, limiting the colonisation or dispersion of the species outside the range distribution of existing adult individuals within the forest. Therefore, a low number of adults trees would result in a low number of seedlings. Personal observations are also congruent with this result, where most of the seedlings were observed under the mature conifer trees (female) than in further areas. This might explain one aspect of the lack of seedlings at the northern population Corral de Salas where no more than 50 individuals were observed within the population (mostly dominated by *N. obliqua*). In contrast, the southernmost population Reigolil showed a large number of adults trees of the species and a high level of seedling density.

Secondly, the significant positive association found between the latitudinal gradient with the number of seedlings (greater seedling density to the south) is most likely explained by climate (temperate and water supply). The southern forests of the country manifest much more stable humid conditions than the central part of Chile. Furthermore, cold conditions remain longer in the south over the year than in the northern forests. *Prumnopitys andina* is a shade-tolerant species that is associated with watercourses and cool temperature. Indeed, the germination process is strongly dependent on long-cold exposure and humid conditions (stratification process) (Gardner et al., 2006). Hence, the lack of conditions meeting these requirements might negatively impact on

regeneration.

This result is consistent with what is suggested by Gao et al. (2017) and Bognounou et al. (2010). Both authors propose that the significant interaction of the species and their latitudinal gradient distribution would be the result of the optimal combination of the environmental conditions. Thus areas less exposed to climate variability would be a better host for some species, as it is shown with *Pr. andina* at the southernmost temperate rain forests of Chile.

Ultimately, the levels of anthropogenic pressures within the forest might also be retarding or stopping the regeneration of *Pr. andina*. Even though this study was not focused on statistically correlating the degree of regeneration with anthropogenic alterations of the ecosystems, we noticed that some populations with a high level of threats were always associated with low number of seedlings. For instance, the coastal population Nahuelbuta (with no regeneration) is characterised by the dominance of exotic plantations (*Pinus* and *Eucalyptus*). Commercial plantations usually cause severe modifications to the ecosystems including, the reduction of water supply, soil acidification and soil impoverishment (lack of nutrients). Therefore, these alterations might limit regeneration for this shade-tolerant species, which favours soils rich in nutrients, humid areas and places near to watercourses.

Other frequent threats observed throughout the distribution of *Pr. andina* include the presence of animals (livestock) and illegal logging (not necessarily in *Prumnopitys*), both threats are also associated with a low number of seedlings. This pattern is congruent with other investigations. For instance, La Manna et al. (2008) evaluated the impact of cattle on sapling establishment in a South American conifer (*A. chilensis*). The authors noted that the lack of saplings was strongly associated with the presence of cattle, mainly due to the constant trampling and browsing caused by the animals. However, over this study, we noticed that cows might also have a positive effect on regeneration but only in the germination process. This because livestock may contribute to the seed scarification of the species (when cows eat the fruits), unlike other farm animals (pig, sheep) which damage the embryo when they eat the fruit. Nevertheless, our results are unresolved in this perspective, and additional systematic studies are recommended.

Possibly suitable condition for *Pr. andina* regeneration.

Here I present a list of variables that could influence positively the natural regeneration of the *Pr. andina*.

- High proportion of *Pr. andina* adults within the forest
- High proportion of *Pr. andina* females
- Low species competition
- Proximity to stream and watercourses
- Cold winter (embryo development, stratification)
- Deep and rich soils
- Shade-light conditions
- Free understory
- low level of threats

Based on a comparison of the number of seedlings, I summary below the populations with high and low levels of seedling density.

Examples of highest levels of seedling density

Lleuques: A small fragmented population in the north-central part of the mountains Andes range. Site associated with shade-light conditions, deep fertile soils, humid areas, watercourses, cold winters. The population also shows free understory (in some areas), low species competition (in some areas), a high number of *Pr. andina* adults and possibly a high proportion of *Pr. andina* females in the population.

Molulco: A medium-size population (Table 4.5) at the south of the range. Site associated with shade-light conditions, deep fertile soils, humid areas, watercourses and cold winters. The population shows free understory (in some areas), a high density of *Pr. andina* adults and a high proportion of *Pr. andina* females in the population.

Conguillio: A medium-size fragmented population at the south of the range. A high level of regeneration only in some areas. Site associated with shade-light conditions, deep fertile soils, humid areas, watercourses, cold winters. A free understory and low species competition (in some regions) is also associated with this population. A high density of *Pr. andina* adults.

Reigolil: The largest and the southernmost population of the species. Site associated with shade-light conditions, possibly deep fertile soils, humid areas, watercourses and cold winters. The population is also associated with a free understory, a high density of *Pr. andina* adults and a high proportion of *Pr. andina* females in the population.

Examples of lowest levels of seedling density

Corral del Salas: The northernmost and smallest population of the species. No regeneration observed at all. Site associated with shade-light conditions, possibly deep and fertile soils, humid areas, watercourses and cold winters. No free understory, high level of competition, low number of *Pr. andina* adults and a small proportion of *Pr. andina* females in the population.

Los Punquios: Big population in the north-central part of the Andean range. Site associated with shade-light conditions, possibly deep and fertile soils, humid areas, watercourses and cold winters. No free understory, high level of competition, low density of *Pr. andina* adults (uncertain about the proportion of *Pr. andina* females in the population).

Laja: A small population at the north-central part of the Andean range. The site also associated with shade-light conditions, possibly deep and rich soils, humid areas, watercourses and cold winters. No free understory, high level of competition, low number of *Pr. andina* adults and not clear about the proportion of *Pr. andina* females in the population.

Laja: Big population at the central-north part of the Andean range. Site associated with shade-light conditions, possibly deep and fertile soils, humid areas, watercourses and cold winters. No free understory, low level of competition in some areas, a small number of *Pr. andina* adults per acres and possibly a high proportion of *Pr. andina* females in the population.

Nahuelbuta (Coastal population): No regeneration was observed in this population. This site is a tiny population, surrounded by exotic plantations, poor soil, high level of competition, no free understory, no humidity. This is one of the most threatened populations with a considerable risk of disappearing due to the high level of pressure (exotic plantations).

Pua: A tiny fragmented population at the south-central valley of Chile. Site associated with shade-light conditions, deep and fertile soils, humid areas, watercourses and cold winters. The population also shows free understory in some areas, a high level of competition, a high number of *Pr. andina* adults and possibly a high proportion of *Pr. andina* females in the population.

Table 4.5: Summary of the natural conditions observed in each population. Populations are listed from the north (left) to the south (right).

	Populations									
	Corral de Salas	**Punquios**	**Lleuques**	**Laja**	**Trapa-Trapa**	**Nahuelbuta**	**Pua**	**Conguillio**	**Molulco**	**Reigolil**
Population size (ha)	5.44	103.74	7.59	5.9	100.97	unknown	3.97	16.22	31.17	301
Shade-light conditions	yes	yes	yes	yes	yes	no	yes	yes	yes	yes
Deep and rich soils	possibly	possibly	yes	possibly	possibly	no	yes	yes	yes	possibly
Humid- areas	yes	yes	yes	yes	yes	no	yes	yes	yes	yes
Associated to watercourses	yes	yes	yes	yes	yes	no	yes	yes	yes	yes
Cold winter	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Free understory	no	no	some areas	no	no	no	some areas	some areas	some areas	yes
Low species competition	no	no	some areas	no	some areas	no	no	some areas	some areas	some areas
Hight n^{o} of *Pr. andina* adults	no	no	yes	no	no	no	yes	yes	yes	yes
High proportion of *Pr. andina* females	no	possibly	possibly	possibly	possibly	no	possibly	possibly	yes	yes
High level of *Pr. andina* regeneration	no	no	yes	no	no	no	no	yes	yes	yes

Natural regeneration in *S. conspicua*, *P. salignus* and *F. cupressides*

To complement this study, informal observations on the regeneration in three other Chilean conifer species were recorded, namely; *S. conspicua*, *P. salignus* and *F. cupressides*. Regeneration (seedlings and saplings) was frequent in the three conifer species throughout their natural distribution. In particular, *P. salignus* was the species which seemed to have the highest level of regeneration (seedlings and saplings). *Fitzroya cupressides* also showed evidence of seedlings in most populations, including a high level of sapling presence. This contrasts, with early studies of the species that concluded little regeneration of the conifer under diverse forest circumstances such as old-growth forest and harvested forest (e.g. Veblen et al. 1976; Donoso et al. 1993; Smith-Ramírez 2007). However, *Fitzroya* seedling frequencies and sapling density were also shown to be positively associated with open canopy and after fire events (Lara et al., 1999). *Saxegothaea conspicua* also showed evidence of regeneration. This result is also congruent with Lusk (1996) who also summarised observation of *S. conspicua* with a continuous regeneration throughout its natural distribution.

4.5 Conclusion

Overall, *Pr. andina* showed frequent seedlings across almost its entire natural distribution. However, a lack of saplings was observed in most populations.

The number of seedling of *Pr. andina* was strongly associated with the number of adult trees and its latitudinal gradient. In contrast, no clear statistical association was detected between the level of seedling competition (from other species) and light intensity. However, light-shade conditions were frequent in all stand-populations.

Several other environmental conditions were also frequent within the *Pr. andina* forests, such as humid areas, watercourses and cold winters and possible deep and fertile soil. However, the most significant factors involved with a higher level of seedling density of the species was the presence of a large number of *Pr. andina* adults in each population. In contrast, there was some evidence for negative impacts on populations from anthropogenic activities.

Two of the ten populations involved in this investigation could be considered under severe risk; the northern population Corral de Salas and the coastal population Nahuelbuta. This mainly due to the low presence of *Pr. andina* adults, the lack of replacement of new individuals (lack of seedlings and sampling) and the high level of anthropogenic pressures from animals and exotic plantations.

Bibliography

Beveridge AE. 1973. Regeneration of podocarps in a central North Island forest. *New Zealand Journal of Forestry*. 18:23–35.

Bognounou F, Tigabu M, Savadogo P, Thiombiano A, Boussim IJ, Oden PC, Guinko S. 2010. Regeneration of five Combretaceae species along a latitudinal gradient in Sahelo-Sudanian zone of Burkina Faso. *Annals of Forest Science*. 67.

Borg H, Stoneman GL, Ward CG. 1988. The effect of logging and regeneration on groundwater, streamflow and stream salinity in the southern forest of Western Australia. *Journal of Hydrology*. 99:253–270.

Carswell FE, Richardson SJ, Doherty JE, Allen RB, Wiser SK. 2007. Where do conifers regenerate after selective harvest?. A case study from a New Zealand conifer-angiosperm forest. *Forest Ecology and Management*. 253:138–147.

Coomes DA, Allen RB, Bentley WA, et al. (13 co-authors). 2005. The hare, the tortoise and the crocodile: The ecology of angiosperm dominance, conifer persistence and fern filtering. *Journal of Ecology*. 93:918–935.

Donoso C, Sandoval V, Grez R, Rodriguez J. 1993. Dynamics of *Fitzroya cupressoides* forests in southern Chile. *Journal of Vegetation Science*. 4:303.

Donoso PJ, Nyland RD. 2005. Seedling density according to structure, dominance and understory cover in old-growth forest stands of the evergreen forest type in the coastal range of Chile. *Revista Chilena de Historia Natural*. 78:51–63.

Donoso Zegers C. 2006. Las Especies Arbóreas de los Bosques Templados de Chile y Argentina : Autoecología. Valdivia: Maria Cuneo, primera edition.

Gao WQ, Ni YY, Xue ZM, Wang XF, Kang FF, Hu J, Gao ZH, Jiang ZP, Liu JF. 2017. Population structure and regeneration dynamics of *Quercus variabilis* along latitudinal and longitudinal gradients. *Ecosphere*. 8.

Gardner M. 2013. *Prumnopitys andina*. The IUCN Red List of Threatened Species

2013.https://dx.doi.org/10.2305/IUCN.UK.2013-1.RLTS.T35934A2861427.en. Downloaded on 27 March 2020.

Gardner MF, Hechenleitner V P, Thomas P, Echeverria C, Escobar B, Brownless P, Martínez C. 2006. Threatened Plants of Central and South Chile: Distribution, Conservation and Propagation. Universidad Austral de Chile and Royal Botanic Garden Edinburgh. 188 pp.

La Manna L, Collantes M, Bava J, Rajchenberg M. 2008. Seedling recruitment of *Austrocedrus chilensis* in relation to cattle use, microsite environment and forest disease. *Ecologia Austral*. 18:27–41.

Lara A, Fraver S, Aravena J, Wolodarsky-Franke A. 1999. Fire and the dynamics of *Fitzroya cupressoides* (Alerce) forests of Chile's Cordillera Pelada. *Ecoscience*. 6:100–109.

Lusk CH. 1996. Stand dynamics of the shade-tolerant conifers *Podocarpus nubigena* and *Saxegothaea conspicua* in Chilean temperate rain forest. *Journal of Vegetation Science*. 7:549–558.

Madsen P. 1995. Effects of soil water content, fertilization, light, weed competition and seedbed type on natural regeneration of beech (*Fagus sylvatica*). *Forest Ecology and Management*. 72:251–264.

Mallik AU. 2003. Conifer regeneration problems in boreal and temperate forests with ericaceous understory: Role of disturbance, seedbed limitation, and keytsone species change. *Critical Reviews in Plant Sciences*. 22:341–366.

Matías L, Jump AS. 2015. Asymmetric changes of growth and reproductive investment herald altitudinal and latitudinal range shifts of two woody species. *Global Change Biology*. 21:882–896.

Oldén A, Komonen A, Tervonen K, Halme P. 2017. Grazing and abandonment determine different tree dynamics in wood-pastures. *Ambio*. 46:227–236.

Parker T, Donoso C. 1993. Natural regeneration of *Fitzroya cupressoides* in Chile and Argentina. *Forest Ecology and Management*. 59:63–85.

Shen C, Nelson AS. 2018. Natural conifer regeneration patterns in temperate forests across the Inland Northwest, USA. *Annals of Forest Science*. 75.

Smith-Ramírez C. 2007. Regeneration of *Fitzroya cupressoides* after indigenous and non-indigenous timber harvesting in southern Chilean forests. *Forest Ecology and Management*. 248:193–201.

Veblen TT, Delmastro RJ, Schlatter JE. 1976. The conservation of *Fitzroya cupressoides* and its environment in southern Chile. *Environmental Conservation*. 3:291–301.

Walck JL, Hidayati SN, Dixon KW, Thompson K, Poschlod P. 2011. Climate change and plant regeneration from seed. *Global Change Biology*. 17:2145–2161.

Wei X, Wu H, Meng H, Pang C, Jiang M. 2015. Regeneration dynamics of *Euptelea pleiospermum* along latitudinal and altitudinal gradients: Trade-offs between seedling and sprout. *Forest Ecology and Management*. 353:232–239.

Wellner CA. 1979. Estimating light intensity beneath coniferous forest canopies : simple field method, volume no.250. Ogden, Utah: Forest Service, U.S. Dept. of Agriculture, Intermountain Forest and Range Experiment Station.

Williams CJ. 2017. Climate Change in Chile: An Analysis of State-of-the-Art Observations, Satellite-Derived Estimates and Climate Model Simulations. *Journal of Earth Science & Climatic Change*. 8:400.

Chapter 5

Conclusion

5.1 Chapters

5.1.1 Overview

This chapter synthesises the main conclusions from all subjects included in this investigation; (a) the optimisation of the *de novo* assembly of loci, using RAD sequencing data, for *Saxegothaea conspicua*, *Prumnopytis andina*, *Podocarpus salignus* and *Fitzroya cupressoides*, (b) a population genetic analysis for the four conifer species and (c) the evaluation of the natural regeneration of *Pr. andina*, including informal notes about the regeneration status of the other three conifer species.

The major threats observed throughout the entire natural distribution of each of the four conifer species will also be discussed in this chapter including a genetic risk assessment, following the protocol of Hollingsworth et al. (2020) Finally, I present a conservation summary and outline potential future work to complement the information given in this thesis.

5.1.2 Summary

Optimisation protocol for the *de novo* assembly of loci.

The main goal of Chapter 2 was focussed on the *de novo* locus assembly from RAD-seq, giving special attention to enhancing our understanding of the impact of parameter value optimisation for these large genome conifer species. Specifically, I assess the optimisation protocol recommended by Paris et al. (2017) and Rochette and Catchen (2017) in four non-model plant species with large genomes and evaluate the sensitivity of parameter settings on locus assembly and population genetic statistics.

Main parameter settings that control the *de novo* assembly using Stacks
The optimised parameter values m (minimum stack depth), M (distance allowed between stacks) and n (distance allowed between catalogue loci) found in each conifer data-set varied between them (except in *S. conspicua* and *P. salignus*) and also differed from the default values defined by Stacks.

A clear impact was observed on the number of polymorphic loci and SNPs retained by using the optimised parameter values. This was most pronounced in the Podocarpaceae species where the number of polymorphic loci and SNPs increased considerably by using the optimised values (compared to the default values). In contrast, there was only a slight effect observed on population genetic inferences, as using the optimised parameter values (with respect to the default values), F_{ST} and pi, only varied lightly.

On the other hand, the level of missing data allowed in each data set impacted considerably the number of polymorphic loci and SNPs retained in each species, with both increasing when allowing a higher level of missing data. In addition, the amount of missing data also impacted on measures of differentiation depending on the software used. Prior to choosing the diveRsity packages (R packages) to calculate genetic differentiation (F_{ST}), the software packages Genodive and Genpop were also tested. These latter software packages typically led to higher estimates of population differentiation. A possible explanation is related to the way that these programs deal with the missing data. These programs by default filled any empty cells with the average allele frequency observed in the data set. In this scenario, levels of missing data might lead to strong and marked bias in any estimate of genetic structure. This is particularly important in a study such as this, where the sample size per population is small, and thus imputing allele frequencies in the presence of missing data can substantially impact on population genetic parameters.

In conclusion, the optimisation protocol performed in each conifer data set provided an efficient framework for exploring and visualising the data of the *de novo* assembly of loci. The Stacks software and the full optimisation protocol provided enough information and details to reduce biases during the analysis and accurately identify the optimised parameter values for *de novo* assembly and downstream analysis.

Population structure in the conifer species

The levels of nucleotide diversity for the four species ranged between pi= 0.11% and 0.27%. The levels of genetic differentiation ranged between F_{ST}= 0.017 and 0.062. This pattern is similar to that found in conifers from the northern hemisphere which usually show more extensive and continuous distribution than the southern hemisphere conifers.

The low level of differentiation detected in the South American conifers was unexpected given the current fragmented distribution of each conifer species and the natural barriers present in the country (the Coastal and Andes ranges). This might suggest that the low differentiation is likely due to the extreme longevity of individuals of these species delaying genetic drift despite their currently fragmented ranges. Gene flow by pollen and/or seed dispersal may also be important for the high level of connectivity found in each species but seems less likely to explain the genetic relatedness, particularly given the complex topography of Chile.

The species demographic history might have also played an essential role in the current genetic status of the Chilean conifers. Indeed, our results suggest reconsidering the hypothesis of a forest expansion from multiples refugia after the LGM to; (a) a forest expansion from a single refugium or (b) a partial forest reduction during the LGM which might have lead also to multiples refugia but still genetically connected.

Levels of regeneration

Frequent seedling presence was observed in almost the entire natural distribution of *Pr. andina*. However, two of the ten populations under investigation showed no regeneration at all (the northern population at Corral de Salas and the Coastal population Nahuelbuta).

The number of seedlings of *Pr. andina* was strongly associated with the number of adults trees and its latitudinal distribution, with the greatest regeneration in the south of the range. In contrast, I did not detect any statistically significant relationship to the level of seedling competition (from other species) and light intensity. However, light-shade conditions were always present in all stand-populations surveyed. Other frequent environmental conditions include humid areas, watercourses and cold winters and possiblly deep and fertile soil. A lack of saplings was prevalent across the entire natural distribution of the species suggesting that external factors might limit sapling establishment such as livestock grazing.

Observations on *S. conspicua*, *P salignus* and *F. cupressoides* forest, suggest that regeneration is usually common for all four species. However, some species such as *P. salignus* showed a higher level of regeneration than *S. conspicua* and *F. cupressoides*. The evidence of saplings was also more evident in these three conifers than in *Pr. andina*. These results differ from earlier investigations of some species, in particular in *F. cupressoides*, where under diverse forest conditions (e.g. old-grow forest and harvested forest) little regeneration was observed.

5.1.3 Conservation implications

Major threats

Multiple threats with different levels of impact were observed for the four conifer species. This includes: human development (housing and urbanisation, tourism and recreation), agricultural activities (wood and pulp plantations, livestock farming), biological resource use (logging and timber extraction) and environmental change (drought). The

threats with the major impact were those associated with agricultural activities (plantation forestry and livestock management) for all four conifers. Pressures from these activities are widespread and frequent affecting a significant area of many of the conifer populations surveyed

I summarise below the main threats observed and their impact by species including general information about the conservation status according to the IUCN. Full details of the main threats recorded by species and their impact are provided in Appendix D.

Saxegothaea conspicua. The species is listed as Near-threatened (NT) by the IUCN. The main threats (concentrated in the coastal populations) were associated with the conversion of land to exotic plantations (*Pinus* and *Eucalyptus*). Livestock was also present in some populations (30% of the populations surveyed). Where livestock was present it has a high impact affecting a significant area of the populations (trampling and browsing seedlings and saplings).

Prumnopitys andina. The species has been classified as vulnerable (VU) by the IUCN. In the coastal range, where only one remnant population exists, *Prumnopitys* is severely threatened by afforestation with exotic species (*P. radiata* and *E. globulus*). This population has also suffered from livestock grazing. Indeed due to the high pressures, this coastal population could be considered with a high chance of extinction in a short time. Livestock is also a common and constant threat in most of the natural distribution of the species (eating fleshy cones, seedlings and small plants). In addition in the Andes, populations have been severely reduced by the foundations of hydroelectric schemes (dams), volcanic activity (e.g. in the National Park Conguillio) and illegal logging of the species. There is not any conservation action for the species so far. In Chile, the species occurs in only three protected areas (N.P Conguillío, N.P. Tolhuaca and N.R Ñuble) which represents a small proportion of the total population.

Podocarpus salignus. The species is listed as vulnerable (VU) by the IUCN. *Podocarpus salignus* has a naturally limited distribution which combined with the conversion of its habitat to commercial plantations (mostly in the northern part of its range), and the devastating effects of fire has resulted in a dramatic decrease over the

last three generations (60 years), to < 30% of the original cover (Allnut et al., 2001). The species is affected mostly by agricultural activities; wood and pulp plantations and livestock. In Chile, there are just a few populations within protected areas. At the coastal range, the species occurs in protected areas at N.R Los Ruiles (Maule region) and Parque Oncol (Los Ríos region). In the Andes, the species is only within the protected area F.R Malleco (Araucanía region).

Fitzroya cupressoides. Alerce (common name) is categorised as an Endangered (EN) species by the IUCN. *Fitzroya* is also internationally protected by the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES), forbidding its commercialisation in any form. Similarly, Chile declared *F. cupressoides* as a Natural Monument in 1976 and together with Argentina prohibits the logging of the species.

In Chile, where 90% of the populations occur, only 18% of its forests are inside the National System of Protected Wilderness Areas. In contrast, in Argentina, 85% of the forests are inside of the National System of Protected Wilderness Areas (Landrum 2006, Kitzberger et al. 2000). The major recorded threat for the species comes from historical over-exploitation, mainly due to its highly prized wood (Veblen and Ashton, 1982), causing the loss of almost half of the original cover (Lara et al., 2008). Fire and also the conversion of forests to pasture land or for exotic plantations have also recorded as significant threats, mostly in the Central valley and the Coastal range populations (Lara et al. 2003, Veblen et al. 1976). In this investigation, we observed only one population with signals of illegal logging (in the coastal population "Astillero"). However, we were not sure if it was an ongoing threat or was from years ago. Yet, another coastal population was classified with a threat of a high impact (Anchile) due to this small population (less than 1 ha) being surrounded by exotic plantations, limiting the expansion of this population in the future.

Wider conservation summary for Chile

Conservation in South America for many years has been mainly dependent on the expansion of the National Park and Reserves systems, from the first Chilean National Park established in the early 1900s with the creation of the National Park "Malleco" in the year 1907 (Sierralta. et al., 2011).

The Chilean National Parks, since the beginning, has been administrated by the National Forestry Corporation (CONAF) dependent first on the Ministry of Agriculture of Chile and now by the Ministry of Environment. Nowadays, CONAF is responsible for administering more than 100 protected areas in the country, totalling 14.56 million ha, accounting for 18% of the flora. However, more than 84% of the protected area belongs to the austral regions (Aysén and Magallanes), leaving the areas which cover the major biodiversity in Chile (Central Chile) unprotected. In addition to the national parks, there are around 1,65 million ha under private conservation initiatives. However, the majority of these are also concentrated in the south of Chile. Indeed, central Chile has been classified as one of the 25 hotspots of biodiversity areas in the world. These areas are those that contain at least 1,500 endemic vascular plant species ($>$ 0.5% of the world total) and that have lost at least 70% of their original habitat (Poynton et al. 2007, Gardner et al. 2006).

The causes of forest reduction and habitat degradation in Chile are numerous. However, one of the causes with the most significative impact, at least in the coastal forests is the conversion of the land to commercial exotic plantations. This is an activity that arose since the promulgation of law No. 701 in 1974 to promote the forest economy in Chile, causing reduction of the native forest (to date), by more than 2.1 million ha (Nahuelhual et al. 2012, Echeverria et al. 2006).

Going beyond these direct human threats, Chile is one of the countries particularly vulnerable to climate change (Stolpe and Undurraga, 2016), mostly due to having low coastal areas; arid and semi-arid zones; areas susceptible to deforestation or erosion, natural disasters, drought and desertification; highly polluted urban areas, and fragile ecosystems.

Because of this, the country is at risk from an increase in temperature and the decrease of precipitation in several regions (Figure 5.1); therefore, many species may become vulnerable to these drastic changes and dependent on their ability to adapt (Garreaud et al., 2017).

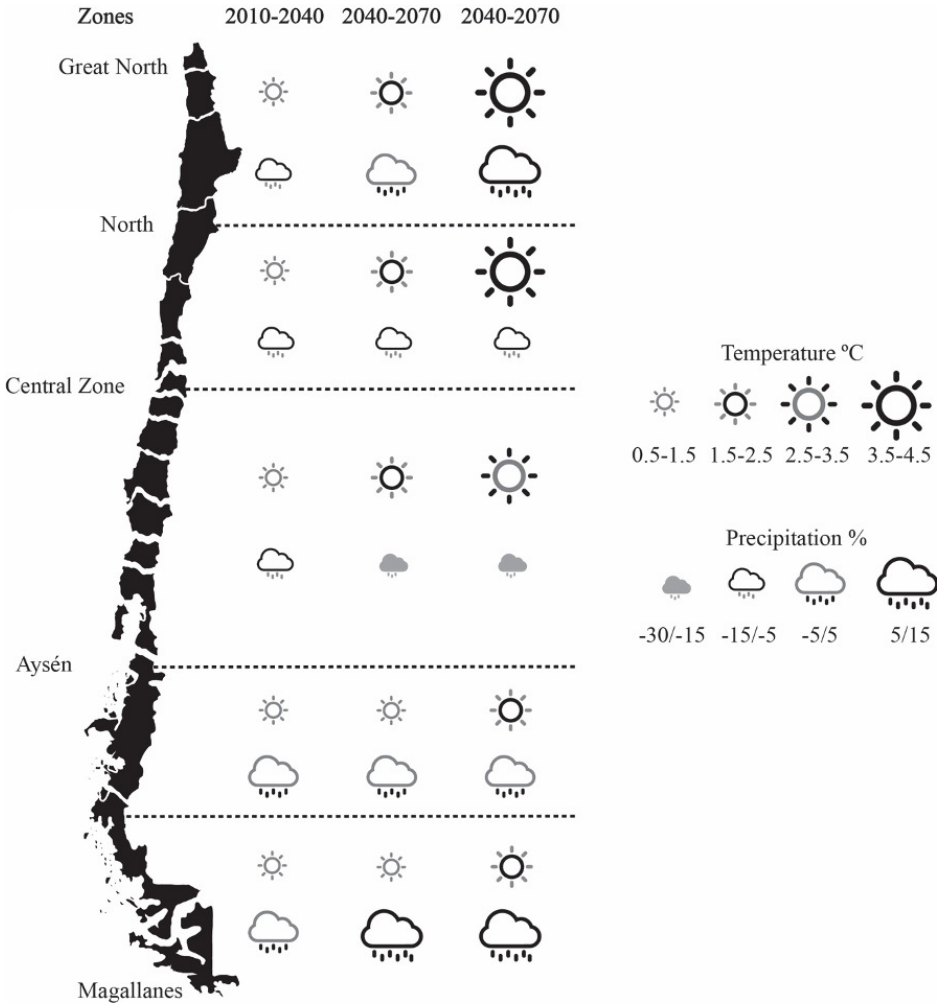


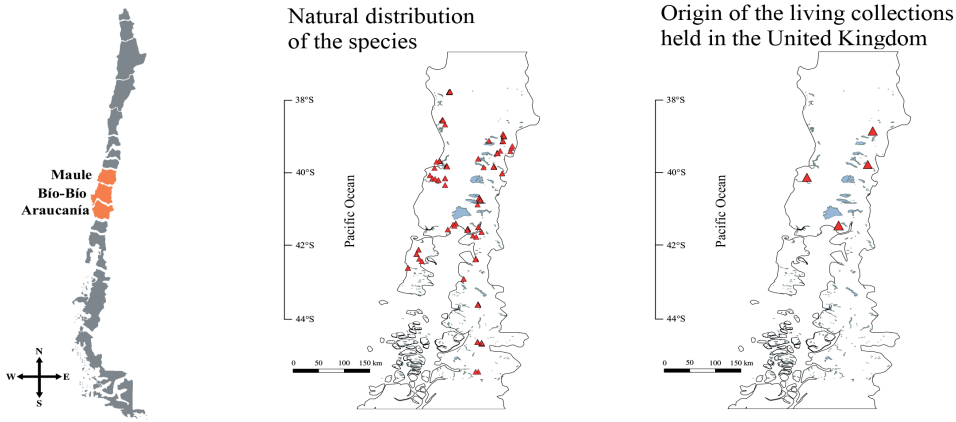
Figure 5.1: Figure taken and modified from The Fifth IPCC Report in its first report (AR5 WG1) "Base Científica del Clima", Terram, December 2013. The figure represents schematically the impacts of climate change according to projections for the future made by ECLAC and associated with the Second National Communication of Chile to the United Nations Framework Convention on Climate Change.

Genetic risk assessment

This thesis has generated information on genetic variation in each of four Chilean conifer species, and also compiled information on regeneration and conservation threats. In this next section, I bring this information together to generate a formal genetic risk assessment for each species. This uses an approach developed by Hollingsworth et al. (2020), which aims to formalise the integration of genetic information into conservation programmes, by providing a standardised template for assessing threats to genetic diversity and evolvability of wild species.

Scientific name	*Saxegothaea conspicua*	**Common Name**	Mañio hembra Prince Albert's Yew
IUCN Category	NT	**T13 Status**	Negligible risk Mitigation not required

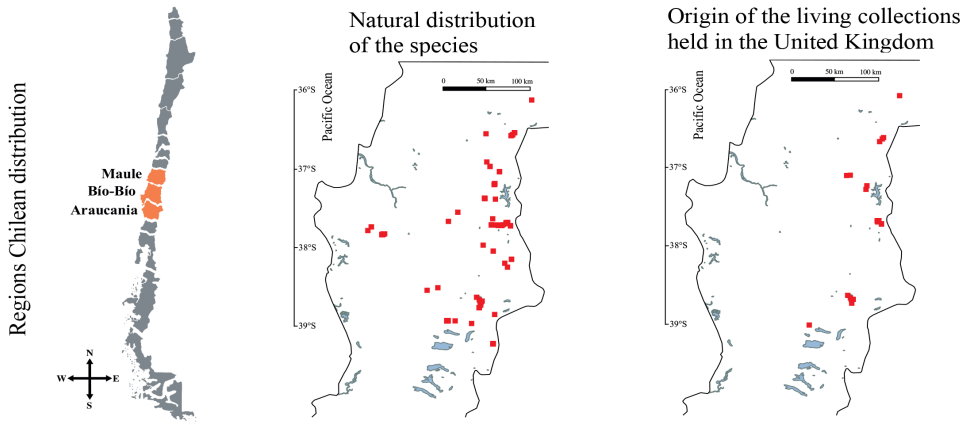
Context	**Background**	Monoecious wind-pollinated species that grows in Chile and Argentina, with most (90%) populations in the temperate forest of Chile (35-46º S). In Chile the species has a relatively continuous distribution (Gardner, 2011). The species produce a cone that usually falls by gravity. No long-distance dispersers are known. Vegetative reproduction seems to be frequent in natural conditions. The species is characterised by having adventitious roots which might extend the longevity of individuals (Cano *et al*., 2014). Molecular studies shown low levels of genetic differentiation (Quiroga, 2009; Cano, 2020). Nuclear Markers (RAD-seq) indicate low levels of genetic differentiation (global F_{ST}= 0.062) and no significant genetic differentiation between populations (Cano *et al*., 2020). Other unpublished investigation using isozyme markers has also observed low levels of genetic differentiation (F_{ST} = 0.082) (Quiroga, 2009).
	Current threats	Agricultural activities represent the major threat of the species (mostly in its coastal distribution) involving conversion of land to exotic plantations (*Pinus* and *Eucalyptus*) and livestock. There are also reports of logging and firewood extraction throughout its range (Gardner 2011).
	Contribution of Chilean population to total species diversity	90% of the species occurrence is within Chile, thus the Chilean populations represent the major stronghold for this species
Genetic risks	**Diversity loss: population declines**	The longevity of individual trees, and its relatively widespread occurrence (at least in its southern distribution) suggests limited risk of imminent genetic diversity loss. However, there is a risk of future genetic diversity loss if existing pressures expand and continue (habitat fragmentation, and reduction of population size caused by the conversion of native forest to commercial plantations and grazing by livestock).

<table>
<tr><td rowspan="4"></td><td>Diversity loss: functional variation</td><td>There is limited risk of loss of adaptive variation, given the apparent range stability (of the southern populations), although the particularly pronounced pressures in the Coastal range may result in a loss of adaptive genetic diversity as the environment in the coastal populations differs from that in the Andean range (and thus underlying adaptive genetic differences are possible).</td></tr>
<tr><td>Diversity loss: divergent lineages</td><td>The species distribution occurs mostly in Chile (90%) thus the loss of the Chilean populations would represent a major loss of the species reservoir of genetic diversity. However, the low levels of population differentiation detected within Chile (Cano, 2020) suggest an absence of divergence intra-specific lineages (and no major genetic differences were noted between Argentinean and Chilean populations using isozymes (Quiroga, 2009).</td></tr>
<tr><td>Hybridisation/ introgression</td><td>Monotypic genus, no hybrids reported.</td></tr>
<tr><td>Low turnover/ constraints on adaptive opportunities</td><td>No evidence of limitation for recruitment of new individuals (Lusk, 1999; Cano et al., 2020). However, heavy grazing can lead to local loss of stands and reduction of seedling establishment (Oldén et al., 2017).</td></tr>
<tr><td rowspan="6">Cumulative risk summary</td><td>In situ genetic threat level</td><td>Negligible (relatively stable range, regeneration widespread, low population differentiation resulting in no imminent threat of genetic diversity loss).</td></tr>
<tr><td>Confidence in in situ threat level</td><td>Medium (assessment based on recent field observations and direct data on genetic variation and population differentiation; no direct data on adaptive variation).</td></tr>
<tr><td>Ex situ representation in the UK</td><td>28 wild collected accessions in the RBGE living collection including a total of 145 plants, distributed in 27 locations in the UK, stemming from four original source populations in Chile.</td></tr>
<tr><td colspan="2"> 
</td></tr>
<tr><td>Current conservation actions</td><td>Important old growth forests protected in the National Parks and private Reserves in the Andean mountains. Ex situ collection established in the UK.
<table>
<tr><td>Ex situ</td><td>Translocation</td><td>Habitat management</td><td>Legal protection of habitat or species</td><td>Control of INNS/pests/ pathogens</td></tr>
<tr><td>X</td><td></td><td></td><td>X</td><td></td></tr>
</table></td></tr>
<tr><td>Overall T13 status</td><td>Negligible, Mitigation not required</td></tr>
</table>

	Overall T13 status explanation	Despite pressures from timber extraction and land-use change the current abundance of the species, the frequent regeneration, and the longevity of individual trees makes imminent loss of genetic diversity unlikely. In the longer term, there is the potential for range contraction (particularly in the northern part of the range) and an associated risk of genetic diversity loss.
	Assessor	Mauricio Cano, Royal Botanic Garden Edinburgh

Scientific name	*Prumnopytis andina*	**Common Name**	Lleuque
IUCN Category	VU	**T13 Status**	Moderate Mitigation not in place

Context	**Background**	Dioecious (rarely monoecious) wind-pollinated tree species. Endemic to Chile with a narrow distribution in the central southern part of the country (35-39º S), with only 12 known locations. The species produces an edible green fruit (with a single seed) which falls by gravity. No long-distance dispersers are known. Molecular studies show low levels of genetic differentiation between populations. Nuclear Markers (RAD-seq) indicate low levels of genetic differentiation (global F_{ST}= 0.060) (Cano *et al*., 2020). Other unpublished studies using isozyme (Quiroga, 2009) and CpDNA markers (Martinez, 2011) also observed low levels of genetic differentiation.
	Current threats	Agricultural activities represent the major threat to the species: conversion of land to plantation forestry (*Pinus* and *Eucalyptus*) and livestock grazing.
	Contribution of Chilean population to total species diversity	Endemic, all genetic diversity in this species is restricted to Chile.
Genetic risks	**Diversity loss: population declines**	Land-use changes are considered to have led to past population declines, and some populations are surrounded by plantation forestry limiting natural expansion. Overall, given the endemic nature of the species, and the small number of extant populations, the loss or decline of any single population would be a proportionally high overall loss for the species. The longevity of individual trees and low levels of population differentiation may limit the loss of genetic diversity, despite the fragmented nature of the species' range. However, long-term progressive genetic erosion remains possible.
	Diversity loss: functional variation	The species occurs in different environments in the Andean and Coastal mountain ranges in Chile, and in the Central Valley. There are no studies on adaptive variation across its range, but there is clear environmental heterogeneity and populations in the southern part of the range experience much more stable humid conditions than the northern populations. Furthermore, the cold conditions remain longer in the south over the year than in the northern forests. Thus, any contraction of the species range may result in loss of adaptive genetic diversity.

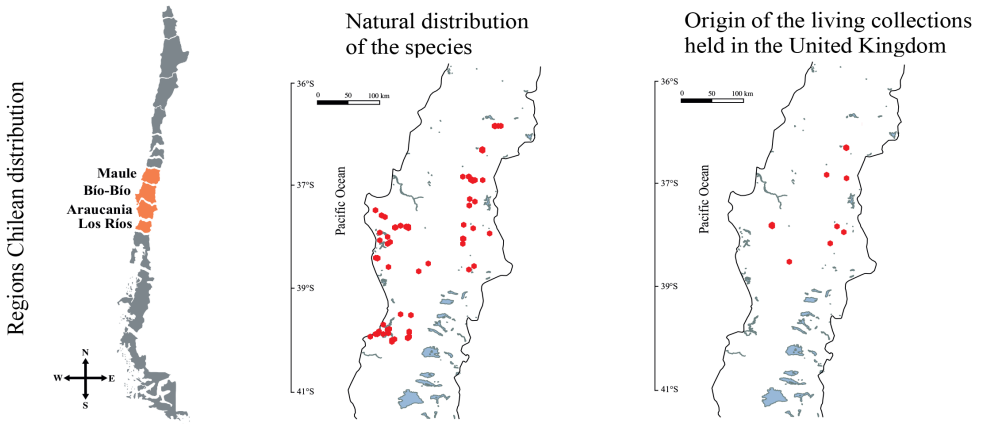
<table>
<tr><td rowspan="3"></td><td>Diversity loss: divergent lineages</td><td>The species is endemic. Its loss from Chile would result in the entire loss of its genetic diversity. Within Chile, there is no evidence of deeply divergent genetic lineages, thus loss of individual populations would not alone result in a threat to intra-specific phylogenetic diversity.</td></tr>
<tr><td>Hybridisation/ introgression</td><td>Low risk – no hybridisation known.</td></tr>
<tr><td>Low turnover/ constraints on adaptive opportunities</td><td>Moderate risk of constraints on regeneration limiting adaptive change. Seedlings are found in most populations, albeit with seedling regeneration more abundant in the south of the range. However, there is a general lack of saplings throughout the distribution of the species (Cano et al., 2020), potentially attributable to livestock grazing of young trees. This may retard adaptive change if new cohorts do not survive to reproductive age</td></tr>
<tr><td rowspan="6">Cumulative risk summary</td><td>In situ genetic threat level</td><td>Moderate (narrow range endemic with extant pressures and lack of sapling establishment, ameliorated by the longevity of individual trees and low levels of population differentiation)</td></tr>
<tr><td>Confidence in in situ threat level</td><td>Medium (assessment based on good demographic data and direct data on genetic variation, population differentiation and biology, further genetic studies with increased sample sizes are needed to better understand patterns of population differentiation and data needed on adaptive differentiation).</td></tr>
<tr><td>Ex situ representation in the UK</td><td>50 wild-collected accessions are managed as part of the living collection of the Royal Botanic Garden Edinburgh (RBGE), including a total of 493 plants from seven populations, cultivated in 31 UK locations. In addition, seeds from the entire natural distribution of the species (from multiple individuals per population) and are currently being germinated at RBGE.</td></tr>
<tr><td colspan="2">


</td></tr>
<tr><td>Current conservation actions</td><td>The species is categorised as Vulnerable by the IUCN and occurs in only three protected areas (Conguillío and Tolhuaca National Park, and the National Reserve Ñuble), representing a small proportion of the total population.
<table>
<tr><td>Ex situ</td><td>Translocation</td><td>Habitat management</td><td>Legal protection of habitat or species</td><td>Control of INNS/pests/ pathogens</td></tr>
<tr><td>X</td><td></td><td></td><td>X</td><td></td></tr>
</table>
</td></tr>
<tr><td>Overall T13 status</td><td>Moderate; Mitigation not in place</td></tr>
</table>

	Overall T13 status explanation	The longevity of individual trees and low levels of detected population differentiation may limit immediate loss of genetic diversity, further supported by extensive *ex situ* conservation. However, the endemic status of the species, its narrow range, and small number of extant populations (most without formal conservation protection) creates an intrinsic set of genetic risks. The absence of saplings may limit adaptive responses to climate change, and the high vulnerability of the only population in the coastal range creates the potential for *in situ* loss of adaptive genetic diversity.
	Assessor	Mauricio Cano, Royal Botanic Garden Edinburgh

Scientific name	*Podocarpus salignus*	**Common Name**	Mañio de hoja larga Willow-leaf Podocarp
IUCN Category	VU	**T13 Status**	Negligible risk Monitoring required

Context	**Background**	Dioecious wind-pollinated bird-dispersed tree species. Endemic to Chile with a narrow distribution in the central southern part of the country (35-40º S). Molecular studies show low levels of genetic differentiation (Allnut *et al*., 1999, Quiroga, 2009, Cano, 2020). Nuclear Markers (RAD-seq) indicate no significant genetic differentiation between populations (global F_{ST} = 0.045) (Cano, 2020). 2n= 38 (probably an ancient polyploid; Cano, 2020).
	Current threats	Conversion of land to exotic plantations (*Pinus* and *Eucalyptus*) and grazing from livestock represent the major threat to the species. Population declines of 30% have recorded over the last 60 years due to fire and conversion of forest to forestry plantations (Gardner 2013).
	Contribution of Chilean population to total species diversity	Endemic – the populations in Chile house the total *in situ* genetic diversity of the species.
Genetic risks	**Diversity loss: population declines**	Risks of immediate loss of genetic diversity considered low due to the low levels of population differentiation, and longevity of individual trees, and relatively widespread extant occurrence. Polyploidy may also help maintain genetic variation, although the nature of polyploidy in this species is poorly understood. Nevertheless, the fragmented nature of its distribution and ongoing reduction of the population sizes due to plantation forestry and grazing creates a risk of longer-term genetic erosion.
	Diversity loss: functional variation	There is limited risk of immediate loss of adaptive variation, as the species range (and the range of environments in which it occurs) is relatively stable. Potential for longer term loss of adaptive variation if range contraction occurs.
	Diversity loss: divergent lineages	Endemic to Chile – loss from Chile would result in complete loss of its phylogenetic diversity. However, within Chile, no divergent intra-specific phylogenetic lineages have been detected thus loss of individual populations would not per se result in loss of phylogenetic diversity.

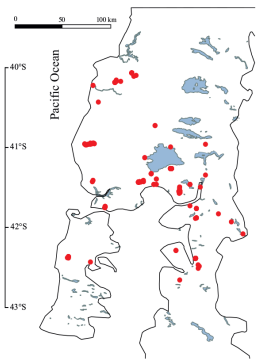
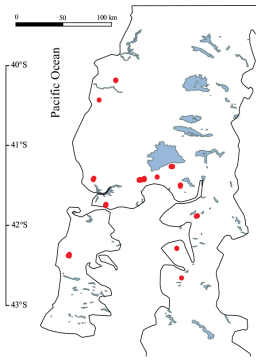
<table>
<tr><td rowspan="1"></td><td>Hybridisation/ introgression</td><td>No hybrids reported. Hybridisation has been reported in ex situ collections with species from New Zealand and this warrants attention for ex situ management (Allnut et al. 2001).</td></tr>
<tr><td></td><td>Low turnover/ constraints on adaptive opportunities</td><td>Informal field observations suggest that regeneration is frequent across the species' range, with no evidence of limitation on recruitment of new individuals (Cano, 2020). However, heavy grazing can lead to local loss of stands and reduction of seedling establishment which may create issues for adaptation in individual populations.</td></tr>
<tr><td rowspan="6">Cumulative risk summary</td><td>In situ genetic threat level</td><td>Negligible (No obvious detectable genetic problems occurring or expected over the next 25 years).</td></tr>
<tr><td>Confidence in in situ threat level</td><td>Medium (assessment based on recent field observations and direct data on genetic variation and population differentiation; no direct data on adaptive variation).</td></tr>
<tr><td>Ex situ representation in the UK</td><td>601 plants from 46 wild collected accessions from seven populations are growing in the living collection managed by the Royal Botanic Garden Edinburgh and distributed over 40 sites in the UK. This material includes populations from the Coastal range and the Andes but does not include representation from the southern part of the species range.</td></tr>
<tr><td colspan="2">


</td></tr>
<tr><td>Current conservation actions</td><td>The species is categorised as Vulnerable by the IUCN. Therefore, its harvesting and any other direct associated threats are restricted, although only three of its populations are located in protected areas (two in the coastal range, and one in the Andes).
<table>
<tr><td>Ex situ</td><td>Translocation</td><td>Habitat management</td><td>Legal protection of habitat or species</td><td>Control of INNS/pests/ pathogens</td></tr>
<tr><td>X</td><td></td><td></td><td>X</td><td></td></tr>
</table>
</td></tr>
<tr><td>Overall T13 status</td><td>Negligible; monitoring required*</td></tr>
<tr><td>Overall T13 status explanation</td><td>Low levels of population differentiation and no evidence for genetic drift among populations, longevity of individual trees, generally good regeneration, representation in ex situ collections. However, the fragmented nature of its distribution, and the steepness of recent populations declines create the possibility of longer-term genetic erosion and review of the status is therefore recommended.</td></tr>
</table>

	Assessor	Mauricio Cano, Royal Botanic Garden Edinburgh
	Other	*The original methodology of Hollingsworth *et al.* (2020) listed four categories to characterise mitigating actions (not required, effective, not effective, not in place). This species highlights the need for a fifth category – 'monitoring required'. I propose the addition of this category to cover species in which the genetic risks over the next 25 years are currently considered negligible, but for which there is a sufficient concern that the situation may deteriorate to warrant a future review of this genetic risk assessment.

Scientific name	*Fitzroya cupressoides*	**Common Name**	Alerce
IUCN Category	EN	**T13 Status**	Negligible Monitoring required

Context	**Background**	Dioecious wind-pollinated tetraploid tree with extreme longevity up to 3620 years (Lara et al., 1993); with seed dispersal via wind. The species grows in Chile and Argentina, with 90% of the populations found in the temperate forest of Chile (39-43ºS). In Argentina it is distributed between the 40-42º S. Molecular studies show low levels of population differentiation (Allnut *et al*., 1999; Cano, 2020). Nuclear Markers (RAD-seq) indicate low levels of no significant population differentiation (global F_{ST}= 0.017) (Cano, 2020). Other studies using RAPD and isozyme markers have also observed low levels of genetic differentiation albeit with the authors inferring evidence for weak regional-scale genetic structure (Allnut *et al*., 1999; Premoli *et al*., 2000b).
	Current threats	Conversion of land to exotic plantations (*Pinus* and *Eucalyptus*) and livestock grazing represent the main extant threat to this species. Historical illegal logging has also had a substantial impact on this species. Large scale fires have also resulted in major impacts of *Fitzroya* populations (Premoli *et al.* 2013).
	Contribution of Chilean population to total species diversity	As 90% of the species range is within Chile, the Chilean populations represents an important component of the species' total genetic diversity. Evidence for weak genetic structure infers differentiation between populations in the west and east of the Andes, and thus some differentiation of the populations in Chile from those in Argentina (Allnut *et al*., 1999; Premoli *et al*., 2000b).
Genetic risks	**Diversity loss: population declines**	The species range at present is stable, and *Fitzroya* is protected across its range as a National Monument. The extreme longevity of trees provides an extremely effective buffer against genetic diversity loss, and genetic drift is likely to be very slow. Nevertheless, there is a pervasive threat from agricultural expansion and catastrophic population loss from large scale human induced wildfires which could lead to longer-term genetic erosion.

	Diversity loss: functional variation	There is likely to be adaptive differentiation in different parts of its range, reflecting the markedly different environmental conditions. For instance, studies have shown clear differences in water use between populations in the Andes and the Coastal ranges which may reflect underlying adaptive differences (Urritia-Jllabert *et al*., 2015). Although the range distribution is currently relatively stable, any future major range contraction would likely lead to loss of adaptive variation.
	Diversity loss: divergent lineages	The species distribution occurs mostly in Chile (90%) and a small fraction in Argentina (10%). Some authors conclude that there is weak regional genetic structure between populations east and west of the Andes, and also between populations in the Coastal Range and the Central Valley (Allnut *et al*., 1999; Premoli *et al*., 2000b). However, all studies to date have shown only weak genetic structure, suggesting little evidence for the prescence of divergent intra-specific lineages, thus likelihood of loss of intra-specific phylogenetic lineages is low.
	Hybridisation/ introgression	Monotypic genus – no hybrids reported or considered likely.
	Low turnover/ constraints on adaptive opportunities	The regeneration of this species has been studied in detail by numerous authors (e.g. Donoso *et al.,* 1993, Lara *et al*., 1999, Smith-Ramirez 2007) and these studies coupled with recent field observations (Cano, 2020) suggest that although at times patchy, there is overall no clear evidence for limitations to regeneration being a major issue for the species.
Cumulative risk summary	***In situ* genetic threat level**	Negligible (no obvious detectable genetic problems expected over the next 25 years).
	Confidence in *in situ* threat level	Medium (assessment based on good demographic data, field observations and direct data on genetic variation; no direct data on adaptive differences between populations).
	***Ex situ* representation in the UK**	117 plants from 48 wild collected accessions from nine populations held in the *ex situ* collection managed by the Royal Botanic Garden Edinburgh (RBGE) distributed across 38 locations in the UK.
	Regions Chilean distribution 	Natural distribution of the species  Origin of the living collections held in the United Kingdom 

<table>
<tr><td></td><td>Current conservation actions</td><td colspan="5">Categorised as Endangered by the IUCN and afforded species-level protection across its range. Approximately 17% of its populations are in protected areas of the Chilean State, and about 20% in private reserves in Chile (Smith-Ramirez et al., 2005). The ex situ collection managed by RBGE encompasses good representation of the species range.
<table>
<tr><td>Ex situ</td><td>Translocation</td><td>Habitat management</td><td>Legal protection of habitat or species</td><td>Control of INNS/pests/ pathogens</td></tr>
<tr><td>X</td><td></td><td></td><td>X</td><td></td></tr>
</table></td></tr>
<tr><td></td><td>Overall T13 status</td><td colspan="5">Negligible; monitoring required*</td></tr>
<tr><td></td><td>Overall T13 status explanation</td><td colspan="5">Although the longevity of individual trees should reduce the likelihood of immediate loss of genetic diversity (potentially further helped by polyploidy), the currently fragmented range, coupled with the pervasive impacts from grazing and the potential for major fires warrants continued monitoring of the genetic status of this iconic species.</td></tr>
<tr><td></td><td>Assessor</td><td colspan="5">Mauricio Cano, Royal Botanic Garden Edinburgh</td></tr>
<tr><td></td><td>Other</td><td colspan="5">*The original methodology of Hollingsworth et al. (2020) listed four categories to characterise mitigating actions (not required, effective, not effective, not in place). This species highlights the need for a fifth category – ‘monitoring required’. I propose the addition of this category to cover species in which the genetic risks over the next 25 years are currently considered negligible, but for which there is a sufficient concern that the situation may deteriorate to warrant a future review of this genetic risk assessment.</td></tr>
</table>

5.2 Future work

Priority areas for future work include a more comprehensive characterisation of conifer population dynamics, more detailed insights into genomic diversity and adaptation, and increased information on environmental pressures threatening conifers to provide essential information for conservation. Below, I discuss some of these key areas for future study.

A highly desirable immediate next step would be generated RAD-seq data from more individuals per population to see if increased individual level sampling provides any further insights into patterns of population genetic structure. Beyond this, it would be advantageous to explore other sequencing approaches for tackling conifer conservation genetics. Even though RAD-seq is one of the most widespread methods for obtaining genomic data from non-model organisms (particularly for populations genetic analyses), this technique is not easily comparable across species, as any differences among data sets in the restriction enzymes or the assembly strategies will reduce the chance of comparability. In contrast, hybrid capture approaches (Suchan et al., 2016) which can recover data from hundreds of nuclear loci, can target a specific set of loci across conifers, thus increasing comparability among studies. At the time of writing, work is underway to produce conifer-specific bait sets (Hollingsworth, pers comm March 2020). Deploying such approaches to conifer population samples has the potential to provide new standardised insights into population genetic structure, and to enable comparative analyses of topics such as the demographic history of populations in different species with similar ranges.

As sequencing technologies improve, it will also be possible for researchers to sequence entire nuclear genomes from many conifer species. This will allow us to find the causal genes involved in adaptation, study the selection pressures shaping genomic diversity, and quantify changes in effective population size over time. Large scale genome sequencing of wild species is still in its infancy, but projects such as the Earth Biogenome Project (Lewin et al., 2018) are leading to major efforts to improve protocols for high molecular weight DNA extractions (necessary for high-quality assembly), and the asso-

ciated informatics challenges of large scale *de novo* assembly of reference genomes.

Given the longevity of conifer species, there is also the potential to combine studies of genetic structure and dendrochronology (Housset et al., 2018). Species such as *Fitzroya cupressoides*, with individuals that exceed 3000 years in age, will have individuals in populations that have established in very different sets of environmental conditions. With high resolution genetic marker data, knowledge of age-structure and past environmental conditions, association studies are possible to explore how variation in functional genes is correlated with the timing (and inferred environmental conditions) of individual establishment and growth. Such approaches ultimately have the potential to add a new dimension to ecological studies of age structure and regeneration.

Another area that could prove useful is the long-term monitoring and revisiting of individual sites. Continuous and systematic survey studies would be very useful for understanding and monitoring the conservation status of the species over time. For example, evaluating the regeneration of *Pr. andina* over time, with special attention to sapling establishment (due to the lack of saplings observed over this investigation). This could provide further information on constraints to the establishment and growth of new individuals within the forest under different levels of threat and also under different changes in temperature and / or water supply. It would also be very useful to include empirical data about the natural regeneration for the other three conifer species studied in this investigation. This could be conducted using a similar methodology to that deployed for *Pr. andina*. For all species, it would also be very useful to include permanent plots to evaluate the levels of regeneration systematically over time. This is extremely useful for the conservation of the species in order to observe the capacity of the conifers to establish new individuals within the forest, which is ultimately critical for the long-term existence of the species.

Assessing the degree of local adaptation of these conifers in a common garden experiment would also be informative. This will provide insights into the levels of adaptive differences among populations, which is of particular importance given the expected changes in environmental conditions due to climate change, and also guide any translocation projects aiming to establish new populations. However, a study of this nature

including Chilean conifer species is challenging due to the difficulties in the germination process of these species and also their long-life cycle. For example, *Fitzroya* shows a very low rate of seed viability (no more than 20%), and only produces seeds every five or seven years. *Prumnopitys* seeds are hard to germinate, needing long cold exposure and scarification. As part of my PhD research, I attempted to germinate and grow seeds of *Prumnopitys* in a common garden site at the RBGE. I established a large-scale common garden experiment of 50 seeds, from 10 mother trees, from 10 populations arranged in a randomised block experiment, with all seeds subjected to stratification (120 days) and scarification to promote regeneration. However, germination was slow after year one and is only occurring now. This material was thus not available for use in this thesis. However, there is now a large collection of experimental plant material that could be used for assessing local adaptation in this species. Further work is required to develop horticultural protocols for *Saxegothaea* germination. Germination in *Podocarpus salignus* is relatively easy– and although there would be challenges in obtaining seed from across its range in a single field season, this species is well suited to common garden experiments.

Bibliography

Allnut TR, Courtis JR, Gardner M, Newton AC. 2001. Genetic variation in wild Chilean and cultivated british populations of *Podocarpus Salignus* D. Don (Podocarpaceae). *Edinburgh Journal of Botany*. 58:459–473.

Cano MJ. 2020. Genetics and Conservation of the Southern South American Conifers. Ph.D. thesis, University of Edinburgh.

Echeverria C, Coomes D, Salas J, Rey-Benayas JM, Lara A, Newton A. 2006. Rapid deforestation and fragmentation of Chilean Temperate Forests. *Biological Conservation*. 130:481–494.

Gardner MF, Hechenleitner PV, Thomas PI, Echeverria C, Escobar B, Brownless P, Martinez CA. 2006. Threatened Plants of Central and South Chile. Universidad Austral de Chile & Royal Botanic Garden Edinburgh., first engl edition.

Garreaud RD, Alvarez-Garreton C, Barichivich J, Pablo Boisier J, Christie D, Galleguillos M, LeQuesne C, McPhee J, Zambrano-Bigiarini M. 2017. The 2010-2015 megadrought in central Chile: Impacts on regional hydroclimate and vegetation. *Hydrology and Earth System Sciences*. 21:6307–6327.

Hollingsworth P, O'Brien D, Ennos R, et al. (43 co-authors). 2020. Scotland's Biodiversity Progress to 2020 Aichi Targets: Aichi Target 13 – Genetic Diversity Maintained – Supplementary Report 2020. Inverness, Scottish Natural Heritage. Technical report.

Housset JM, Nadeau S, Isabel N, Depardieu C, Duchesne I, Lenz P, Girardin MP. 2018. Tree rings provide a new class of phenotypes for genetic associations that foster insights into adaptation of conifers to climate change. *New Phytologist*. 218:630–645.

Kitzberger T, Perez A, Iglesias G, Premoli AC, Veblen TT. 2000. Distribución y estado de conservación del alerce (*Fitzroya cupressoides* (Mol.) Johnst.) en Argentina. *Bosque*. 21:79–89.

Landrum LR. 2006. Variación intraespecifica en las especies arbóreas de los bosques templados de Chile y Argentina. *Economic Botany*. 60:198–198.

Lara A, Echeverría C, Thiers O, Huss E, Escobar B, Tripp K, Zamorano C, Altamirano A. 2008. Restauración ecológica de coníferas longevas: el caso del alerce (*Fitzroya cupressoides*) en el sur de Chile. *Restauración de bosques en América Latina*. pp. Páginas 39–56.

Lara A, Gardner M, Vergara R. 2003. The use and conservation of *Fitzroya cupressoides* (Alerce) forests in Chile. *Acta horticulturae*. 615:381–385.

Lewin HA, Robinson GE, Kress WJ, et al. (24 co-authors). 2018. Earth BioGenome Project: Sequencing life for the future of life. *Proceedings of the National Academy of Sciences of the United States of America*. 115:4325–4333.

Nahuelhual L, Carmona A, Lara A, Echeverría C, González ME. 2012. Land-cover change to forest plantations: Proximate causes and implications for the landscape in south-central Chile. *Landscape and Urban Planning*. 107:12–20.

Paris JR, Stevens JR, Catchen JM, Johnston S. 2017. Lost in parameter space: a road map for stacks. *Methods in Ecology and Evolution*. 8:1360–1373.

Poynton JC, Loader SP, Sherratt E, Clarke BT. 2007. Biodiversity hotspots for conservation priorities. *Biodiversity and Conservation*. 16:853–858.

Rochette NC, Catchen JM. 2017. Deriving genotypes from RAD-seq short-read data using Stacks. *Nature Protocols*. 12:2640–2659.

Sierralta L, Serrano R, Rovira J, Cortés C. 2011. Las áreas protegidas de Chile. Technical report, Ministerio del Medio Ambiente, Santiago, Chile.

Stolpe N, Undurraga P. 2016. Long term climatic trends in Chile and effects on soil moisture and temperature regimes. *Chilean Journal of Agricultural Research*. 76:487–496.

Suchan T, Pitteloud C, Gerasimova NS, Kostikova A, Schmid S, Arrigo N, Pajkovic M, Ronikier M, Alvarez N. 2016. Hybridization capture using RAD probes (hyRAD), a new tool for performing genomic analyses on collection specimens. *PLoS ONE*. 11:1–22.

Veblen T, Ashton D. 1982. The regeneration status of *Fitzroya cupressoides* in the Cordillera Pelada, Chile. *Biological Conservation*. 23:141.

Veblen TT, Delmastro RJ, Schlatter JE. 1976. The Conservation of *Fitzroya cupressoides* and its environment in southern Chile. *Environmental Conservation*. 3:291–301.

Appendix A

Appendix from Chapter 2

DNA extraction

Here I present the full DNA extraction protocol using the Plant DNeasy Kit (QIAGEN). *Note Buffer AP1 concentrate may form precipitates upon storage. If necessary, warm to 65 ℃ in a water-bath to redissolve.*

1. Add 400 µL Buffer AP1 and 4 µL RNase a stock solution (100 mg/ml) to a maximum of 100 mg (wet weight) or 20 mg (dried) disrupted plant or fungal tissue and vortex vigorously, no tissues clumps should be visible. Vortex or pipet further to remove any clumps of tissue properly and will, therefore, result in a lower yield of DNA.

2. Incubate the mixture for 1 hour at 65 ℃. Mix 2-3 times during incubation by inverting tube, or use a Thermomixer set at 800 rpm. This step lyses the cells. The manufacturer suggests a lysis step of 10 mins.

3. Add 130 µL Buffer P3 to the lysate, mix, and incubate for 5 min on ice. This step precipitates detergent, proteins, and polysaccharides.

4. Centrifuge the lysate for 5 min at 13,000 rpm. Some plants material can generate very viscous lysates and large amounts of precipitates during this step. This can result in the shearing of DNA in the next step. In this case, optimal results are obtained if the majority of these precipitates are removed by centrifugation for 5 min 13,000 rpm.

5. Pipet the lysate into the QIAshredder Mini spin column (lilac) placed in a 2ml collection tube, and centrifuge for 2 min at 13,000 rpm. It may be necessary to cut the end of the pipet tip to apply lysate to the QIAshredder Mini spin column. The QIAshredder Mini spin column removes most precipitates and cell debris, but a small amount will pass through and form a pellet in the collection tube. Be careful not to disturb this pellet in step 6.

6. Transfer the flow-through fraction from step 5 into a new collection tube without disturbing the cell-debris pellet. Typically, 450 µL of lysate in recovered.

7. Add 650 µL of Buffer AW1 to the cleared lysate and mix by pipetting. Using the same pipette tip, transfer 650 µL of the mixture, including any precipitate that may have formed, into the DNeasy Mini spin column placed in a 2 ml collection tube (supplied).

8. Centrifuge for 1 min at 8000 rpm and discard the flow-through into the appropriate waste bottle.

9. Transfer the remaining samples into the DNeasy Mini spin column and repeat step 8. Discard the flow-through into the appropriate waste bottle and discard the collection tube.

10. Place the DNeasy Mini spin column into a new 2 ml collection tube and add 500 µL Buffer AW2, and centrifuge for 1 min at 8000 rpm. Discard the flow-through and reuse the collection tube in step 11. Add 500 l Buffer AW2 to the DNeasy Mini spin column, and centrifuge for 2 min at 13,000 rpm to dry the membrane. It is essential to drain the membrane of the DNeasy Mini spin column since residual ethanol may interfere with subsequent reactions. This centrifugation step ensures that no residual ethanol will be carried over elution. Discard the flow-through and spin again at 13,0000 rpm for 30 seconds to ensure the membrane is dehydrated. Discard the collection tube.

11. Transfer the DNeasy Mini spin column to a new, clean 1.5 ml microcentrifuge tube and pipet 100 µL Buffer AE directly onto the DNeasy membrane. Incubate for 5 min at room temperature (15-25 ℃), and then centrifuge for 1 min at 8000 rpm to elute. Elution with 50 µL (instead of 100 µL) increase the final DNA concentration in the eluate significantly, but also reduce overall DNA yield.

12. Check DNA quality and concentration.

13. Store DNA at -20 ℃.

Table A.1: DNA concentrations obtained for the conifer species (*S. conspicua*, *Pr. andina*, *P. salignus* and *F. cupressoides*) following DNA extraction using the DNeasy^{TM} Plant Minikit.

S. conspicua		*Pr. andina*		*P. salignus*		*F. cupressoides*	
Nº collection	DNA (ng/ul)	Nº collection	DNA (ng/ul)	Nº collection	DNA (ng/ul)	Nº collection	DNA (ng/ul)
SA1_294	96.7	PA1_1	28.2	PS1_37	35	F1_919	26.9
SA1_299	100	PA1_9	26.9	PS1_55	37	F1_933	33.7
SA1_302	102.4	PA1_23	30.1	PS1_57	35	F1_940	46.9
SA1_1	44.4	PA1_13	75.93	PS1_7	87.9	F4_1152	27.2
SA1_2	64	PA1_14	51.4	PS1_8	59.5	F4_1174	33.7
SA2_709	107.4	PA3_126	25.7	PS2_162	33	F4_1176	43.7
SA2_714	99	PA3_133	27.9	PS2_164	31	F5_1179	41.5
SA2_721	93.8	PA3_149	30.3	PS2_170	32	F5_1187	34.8
SA5_834	18.95	PA4_186	29.6	PS4_270	25	F5_1203	47.1
SA5_841	102.1	PA4_192	31	PS4_275	26	F6_1205	34
SA5_845	21.05	PA4_201	33	PS4_277	22.5	F6_1206	33.5
SA6_3	22.5	PA5_245	31	PS6_341	23.6	F6_1215	35.7
SA6_4	21.5	PA5_251	45	PS6_357	27.7	F7_1262	27.8
SA8_963	37.3	PA5_255	34	PS6_362	25.6	F7_1272	35.4

Table A.1 continued from previous page

SA8_965	32.6	PA6_368	34	PS7_393	38	F7_1274	41.4
SA8_974	37.6	PA6_381	34.8	PS7_395	28	F8_1331	34.7
SA8_5	21.5	PA6_387	32.8	PS7_406	31	F8_1337	35.5
SA8_6	16.05	PA6_743	33	PS8_857	26.2	F8_1354	30.4
SA9_1103	40.3	PA7_745	36.3	PS8_873	27	F9_1410	38.4
SA9_1114	30	PA7_755	33.7	PS8_882	27	F9_1418	33.9
SA9_1118	36.4	PA7_761	31.4	PS9_1243	99.6	F9_1430	34.5
SA10_1313	46.4	PA8_770	31.8	PS9_1249	27.4	F10_667	32.2
SA10_1316	43.3	PA8_777	29.7	PS9_1251	23.7	F10_674	26.1
SA10_1322	50.9	PA8_15	61.9	PS9_9	37.2	F10_683	27.4
SA11_632	26	PA8_16	72.53	PS9_10	27.3		
SA11_636	86.9	PA10_1381	34.4	PS11_1356	27.2		
SA11_647	86.3	PA10_1393	33.3	PS11_1361	28		
SA12_489	23.4	PA10_1404	38.7	PS11_1374	26.6		
SA12_506	36.9	PA10_17	56.1	PS11_11	58.4		
		PA10_18	68.5	PS11_12	89.1		

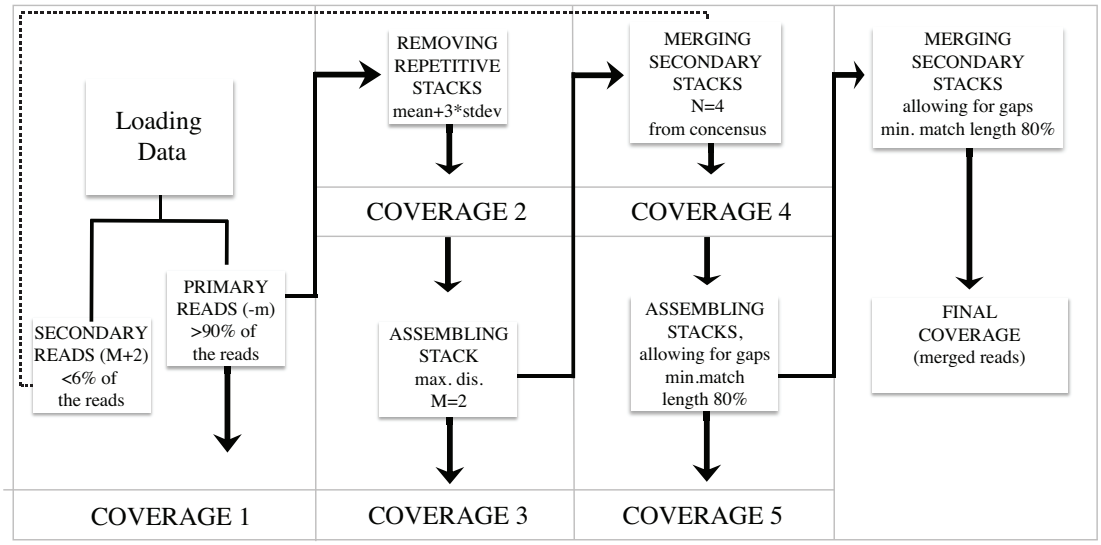


Figure A.1: Coverage layout formation in Stacks. The final coverage formation is controlled by; a) The minimum number of reads (primary reads) to create a stack (m). Reads that did not pass the primary stack formation, generate a group of secondary reads forming secondary stacks, these reads are used later, after the assembling stacks step. b) Removing repetitive stacks. Here the coverage decreases according to the number of repetitive regions present in the data set. c) Maximum distance allowed between stacks (M=2), which affect the stacks assembly. d) Maximum distance allowed to align secondary reads (M+2 or N=4). e) Gapped assembly (80%) for primary stacks and secondary stacks, a filter that allows assembling stacks with a percentage of mismatches. Coverage 1 > Coverage 2; Coverage 2 < Coverage 3; Coverage 3 < Coverage 4; Coverage 4 < Coverage 5 and Coverage 5 < Final Coverage.

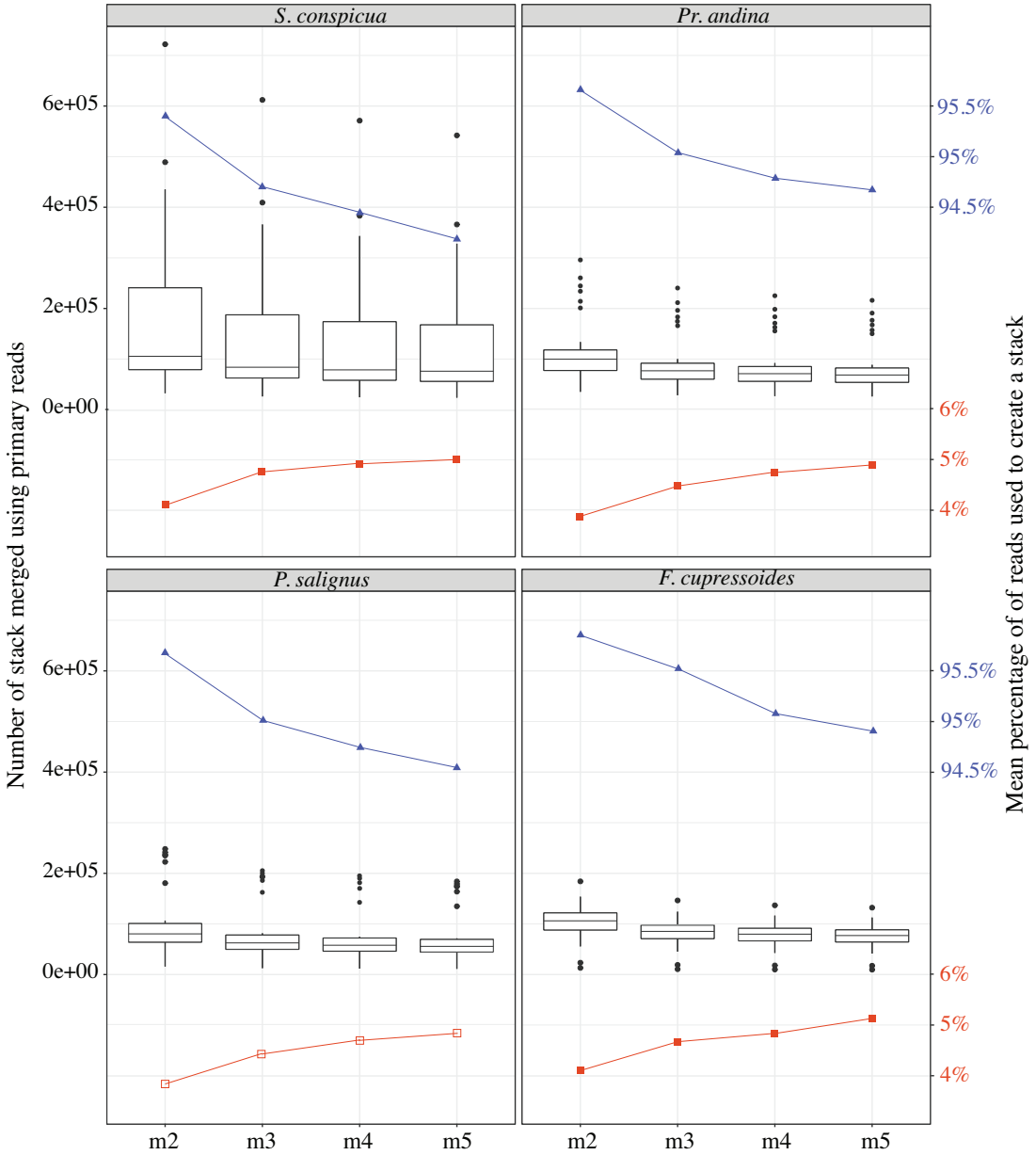


Figure A.2: Distribution of the number of stack formation using primary reads (identical reads to create a stack) and percentage distribution of the primary and secondary reads merged to form a stack by each iteration of m. By default, secondary reads allow gaps equivalent at M+2 (M; distance allowed between stacks). The left Y-axis indicates the number of stacks merged by each iteration of m using primary reads. The right Y-axis represents the mean percentage of the primary and secondary reads used to create a stack. The X-axis indicates the iteration of the minimum number of reads to form a stack (m). The boxplot represents the mean number of stacks created by each iteration of m. The blue-line represent the mean percentage of primary reads used to create a stack and in red-line the mean percentage of secondary reads used to form a secondary-stack

Appendix B

Appendix from Chapter 3

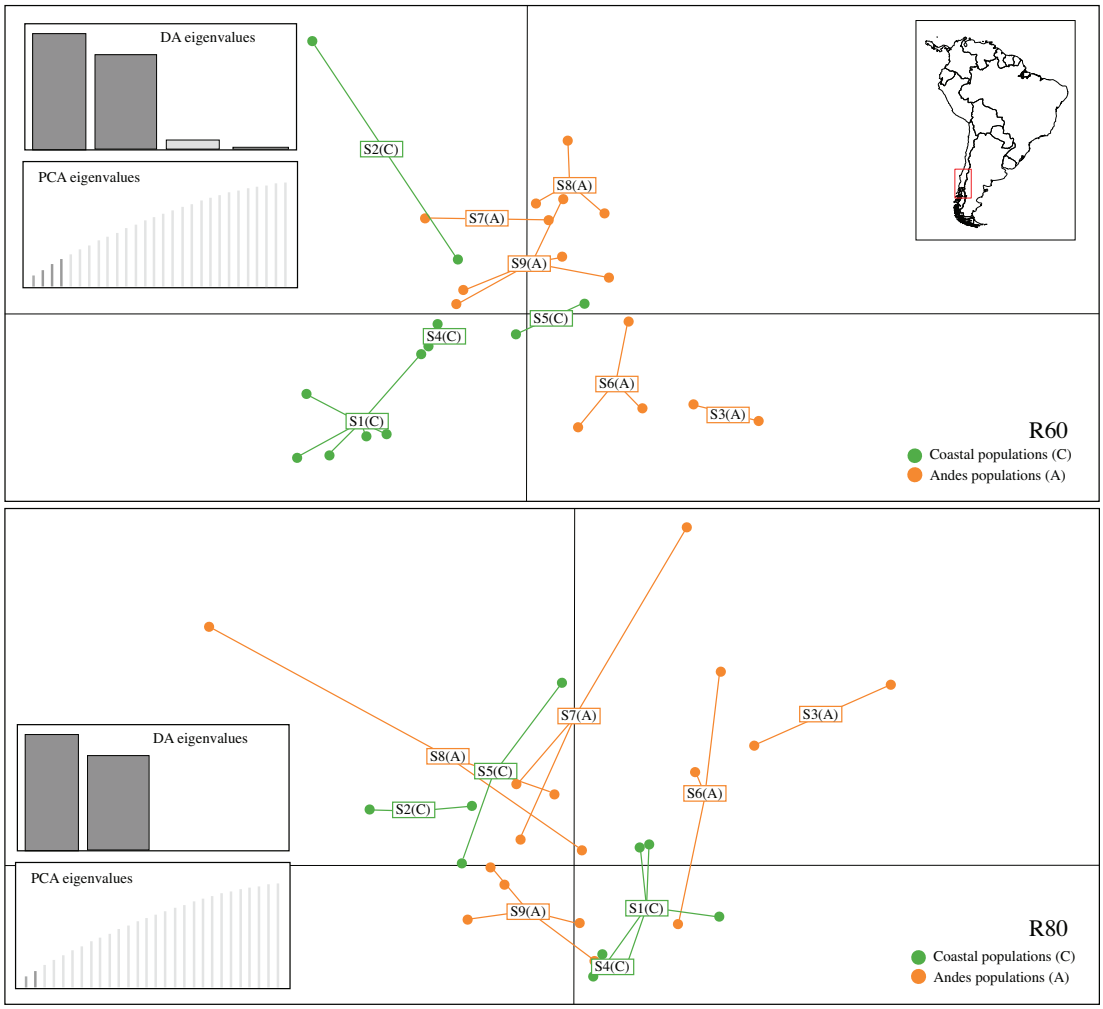


Figure B.1: Impact of changing the missing data threshold on the DAPC analysis. The example here is for one species (*S. conspicua*). However, this pattern is similar for all species. At the top, DAPC assessed with about 40% of missing data (R60). DAPC at the botton with about 20% missing data (R80).

Appendix C

Appendix from Chapter 4

General stand information

Table C.1: List of the largest trees by population. The list of the largest trees in each population is based on the two largest individuals (DBH-height) found in each stand by each population. 2 trees x 10 stands (n= 20) were assessed in most populations except for Corral de Salas which was assessed by 3 stands (total n= 6) and Los Punquios with 9 (total n= 18) stands. Populations are listed from the north to the south and tree species from the largest to the smallest DBH by population.

Corral de Salas (3-stands)		
Species	**DBH**	**Height**
Nothofagus obliqua	60	11
Nothofagus obliqua	50	12
Nothofagus obliqua	40	10
Prumnopitys andina	40	14
Nothofagus obliqua	35	8
Nothofagus obliqua	30	10
Los Punquios (9-stands)		
Austrocedrus chilensis	65	20
Prumnopitys andina	60	15
Prumnopitys andina	50	12
Prumnopitys andina	45	10
Prumnopitys andina	45	13
Prumnopitys andina	40	10
Austrocedrus chilensis	40	15
Prumnopitys andina	40	14
Prumnopitys andina	35	10
Nothofagus obliqua	35	16
Nothofagus obliqua	35	16
Laurelia sempervirens	30	12
Nothofagus obliqua	30	15
Nothofagus obliqua	30	16
Nothofagus obliqua	25	10
Nothofagus obliqua	25	12

Table C.1 continued from previous page

Austrocedrus chilensis	25	12
Prumnopitys andina	20	9
Nothofagus obliqua	20	15
Prumnopitys andina	15	8
Los Lleuques (10-stands)		
Prumnopitys andina	100	12
Prumnopitys andina	90	14
Prumnopitys andina	65	12
Nothofagus obliqua	40	25
Nothofagus dombeyi	40	24
Prumnopitys andina	35	12
Nothofagus obliqua	35	17
Nothofagus dombeyi	35	20
Nothofagus dombeyi	30	18
Nothofagus obliqua	30	20
Nothofagus obliqua	30	22
Nothofagus obliqua	30	20
Nothofagus obliqua	30	15
Prumnopitys andina	25	14
Nothofagus obliqua	20	15
Nothofagus obliqua	20	14
Laja (10-stands)		
Prumnopitys andina	100	18
Nothofagus obliqua	70	23
Nothofagus dombeyi	60	20
Prumnopitys andina	60	13
Quillaja saponaria	50	16
Prumnopitys andina	40	14
Prumnopitys andina	40	15

Table C.1 continued from previous page

Prumnopitys andina	40	12
Nothofagus obliqua	35	14
Nothofagus dombeyi	35	20
Nothofagus obliqua	35	16
Prumnopitys andina	30	14
Prumnopitys andina	25	13
Prumnopitys andina	25	26
Prumnopitys andina	15	8
Nothofagus obliqua	15	14
Nothofagus obliqua	15	14
Quillaja saponaria	10	6
Dasyphyllum diacanthoides	10	6
Prumnopitys andina	5	7
Trapa Trapa (10-stands)		
Prumnopitys andina	90	25
Prumnopitys andina	80	25
Nothofagus dombeyi	80	22
Nothofagus obliqua	60	15
Prumnopitys andina	60	14
Prumnopitys andina	60	14
Prumnopitys andina	60	12
Prumnopitys andina	60	14
Prumnopitys andina	60	16
Prumnopitys andina	50	22
Prumnopitys andina	50	12
Prumnopitys andina	50	11
Prumnopitys andina	50	13
Prumnopitys andina	50	14
Prumnopitys andina	40	10

Table C.1 continued from previous page

Prumnopitys andina	40	14
Prumnopitys andina	35	16
Prumnopitys andina	35	10
Prumnopitys andina	30	8
Prumnopitys andina	20	8
Pua-Santa Lucia (10-stands)		
Nothofagus obliqua	80	26
Nothofagus obliqua	75	25
Prumnopitys andina	70	26
Prumnopitys andina	70	26
Nothofagus obliqua	70	27
Prumnopitys andina	70	2
Persea lingue	60	25
Prumnopitys andina	60	23
Prumnopitys andina	60	24
Prumnopitys andina	60	25
Prumnopitys andina	50	20
Persea lingue	50	24
Prumnopitys andina	45	20
Prumnopitys andina	40	20
Prumnopitys andina	40	24
Persea lingue	35	25
Prumnopitys andina	30	18
Prumnopitys andina	30	20
Prumnopitys andina	30	22
Prumnopitys andina	25	22
Conguillio (10-stands)		
Nothofagus obliqua	80	26
Nothofagus obliqua	70	22

Table C.1 continued from previous page

Nothofagus obliqua	70	25
Prumnopitys andina	70	22
Laurelia sempervirens	60	25
Laurelia sempervirens	50	25
Nothofagus obliqua	50	14
Prumnopitys andina	50	18
Nothofagus obliqua	45	25
Prumnopitys andina	45	14
Nothofagus obliqua	40	26
Laurelia sempervirens	40	18
Prumnopitys andina	40	16
Nothofagus obliqua	40	18
Nothofagus obliqua	35	22
Nothofagus obliqua	35	25
Prumnopitys andina	35	17
Prumnopitys andina	35	15
Nothofagus obliqua	30	20
Nothofagus antartica	20	10
Molulco (10-stands)		
Nothofagus antartica	70	20
Nothofagus dombeyi	60	25
Nothofagus obliqua	60	26
Prumnopitys andina	60	16
Nothofagus obliqua	50	25
Prumnopitys andina	50	17
Prumnopitys andina	50	16
Nothofagus antartica	45	26
Nothofagus obliqua	40	23
Nothofagus obliqua	40	22

Table C.1 continued from previous page

Nothofagus obliqua	40	21
Nothofagus antartica	40	20
Nothofagus obliqua	35	23
Nothofagus obliqua	35	22
Prumnopitys andina	35	12
Nothofagus obliqua	35	18
Nothofagus obliqua	30	20
Nothofagus obliqua	30	20
Nothofagus obliqua	30	18
Nothofagus obliqua	30	25
Reigolil (10-stands)		
Prumnopitys andina	60	16
Prumnopitys andina	60	14
Prumnopitys andina	50	14
Nothofagus dombeyi	50	16
Prumnopitys andina	50	13
Prumnopitys andina	50	14
Nothofagus dombeyi	40	12
Prumnopitys andina	40	12
Prumnopitys andina	40	11
Nothofagus obliqua	40	16
Prumnopitys andina	40	12
Prumnopitys andina	40	12
Prumnopitys andina	35	13
Prumnopitys andina	35	12
Prumnopitys andina	35	14
Prumnopitys andina	30	10
Prumnopitys andina	25	12
Prumnopitys andina	25	10

Table C.2: List of associated species to the *Pr. andina* forest by life form and its percentage of occurrence (respect to the total Plots surveyed). A total of 82 plots were surveyed

Family	Species	Occurrence (nº)	Occurrence. (%)
	Trees		
Podocarpaceae	*Prumnopitys andina*	82	100
Nothofagaceae	*Nothofagus obliqua*	47	57.3
Proteaceae	*Lomatia hirsuta*	32	39
Eleocarpaceae	*Aristotelia chilensis*	18	22
Asteraceae	*Dasyphyllum dicanthoides*	13	15.9
Myrtaceae	*Luma apiculata*	12	14.6
Nothofagaceae	*Nothofagus dombeyi*	12	14.6
Cupressaceae	*Austrocedrus chilensis*	11	13.4
Verbenaceae	*Rhaphithamnus spinosus*	11	13.4
Aextoxicaceae	*Aextoxicon punctatum*	10	12.2
Proteaceae	*Lomatia dentata*	10	12.2
Podocarpaceae	*Podocarpus salignus*	10	12.2
Proteaceae	*Gevuina avellana*	9	11
Lauraceae	*Persea lingue*	7	8.5
Winteraceae	*Drimys winteri*	6	7.3
Nothofagaceae	*Nothofagus antartica*	6	7.3
Fabaceae	*Sophora microphylla*	5	6.1
Celastraceae	*Maytenus boaria*	4	4.9
Proteaceae	*Embothrium coccineum*	3	3.7
Monimaceae	*Laurelia semperverins*	3	3.7
Myrtaceae	*Myrceugenia ovata*	2	2.4
Anacardiaceae	*Schinus patagonicus*	1	1.2
	Shrub		
Flacourtiaceae	*Azara petiolaris*	9	11
Flacourtiaceae	*Azara microphylla*	5	6.1
Fabaceae	*Fabaceae sp*	5	6.1
Ericaceae	*Gaultheria sp*	4	4.9
Rhamnaceae	*Sp*	4	4.9
Rhamnaceae	*Rhamnus diffusus*	3	3.7
Flacourtiaceae	*Azara serrata*	2	2.4
Asteraceae	*Baccharis sp*	2	2.4
Rhamnaceae	*Colletia hystrix*	2	2.4
Flacourtiaceae	*Azara integrifolia*	1	1.2
Berberidaceae	*Berberis microphylla*	1	1.2
Berberidaceae	*Berberis sp*	1	1.2
Santalaceae	*Myoschilo oblonga*	1	1.2
Rosaceae	*Rubus ulmifolius*	1	1.2
Grossulariaceae	*Ribes magellanicum*	1	1.2
	Herb and climbers		
Poaceae	*Chusquea sp*	21	25.6
Hydrangeaceae	*Hydrangea serratifolia*	4	4.9
Boraginaceae	*Heliotropium chilense*	1	1.2
Vitaceae	*Cissus striata*	1	1.2

Table C.3: List of species found across populations and its Midpoint coverage range (%) based on Braun Blanquet cover abundance scale (Coversion of the Braun Blanquet to the Midpoint scale in Table C.5. Full list of the Braun Blanquet scale (previous conversion to the Midpoint scale) for each population in Table C.6). Av. indicate the average of the coverage abundance across populations by species. sp refers to unidentified species

Species/Populations	Corral de Salas	Los Punquios	Los Lleuques	Laja	Trapa-Trapa	Pua	Conguillio	Molulco	Reigolil	Av (%)
				Trees						
Prumnopitys andina	6.7	30.3	20	14.3	85	53	52.8	85	70	41.7
Nothofagus obliqua	79.2	18.3	13.1	3.5	1	6	12.8	0.5	15	14.9
Lomatia hirsuta	0	5.3	7.2	1.8	4.5		2.3		4.5	2.6
Dasyphyllum diacanthoides			6.1	17						2.3
Persea lingue						18.5				1.9
Lomatia dentata				11		4	0.5			1.6
Azara petiolaris		14.5		0.3						1.5
Nothofagus dombeyi			8.1	2	1.8			1.5	0.3	1.4
Podocarpus salignus			10			2				1.2
Luma apiculata		1.5	0.6	7.3		0.3				1
Aextoxicon punctatum	0.1	0.5		8.8						0.9
Rhaphithamnus spinosus				0.5		5.5	1		0.5	0.8
Aristotelia chilensis				2.8		0.5	3.5	0		0.7
Drimys winteri			5.8							0.6
Laurelia semperverins		0.3					5.3			0.6

Table C.3 continued from previous page

Species/Populations	Corral de Salas	Los Punquios	Los Lleuques	Laja	Trapa-Trapa	Pua	Conguillio	Molulco	Reigolil	Av (%)
Nothofagus antartica							3.3		2	0.5
Gevuina avellana				0.8			4			0.5
Sophora microphylla				0.5			3.3			0.4
Embothrium coccineum			3.6							0.4
Austrocedrus chilensis		1.3		0.8	0		0.5			0.3
Maytenus boaria				0.8						0.1
Myrceugenia ovata				0.3					0.3	0.1
Maytenus chilensis				0.3						0
Schinus patagonicus				0.3						0
Shurbs										
Fabaceae sp	13.3	5.5								1.9
sp	5	0								0.5
Gaultheria sp			2.2	1.5						0.4
Rhamnus diffusus						2				0.2
Rubus ulmifolius						1.8				0.2
Ribes magellanicum							1.5			0.2
Berberis microphylla				1.5						0.2
Azara microphylla			0.3	0.3			0.3		0.5	0.1
Azara serrata			0.6							0.1

Table C.3 continued from previous page

Species/Populations	Corral de Salas	Los Punquios	Los Lleuques	Laja	Trapa-Trapa	Pua	Conguillio	Molulco	Reigolil	Av (%)
Baccharis sp		0.3	0.3		0					0.1
Colletia hystrix			0.3	0.3						0.1
Azara integrifolia					0					0
Myoschilo oblonga				0						0
Berberis sp										0
Herbs and climbers										
Chusquea sp		10.8			2	5.8			0.3	1.9
Heliotropium chilense				0.3						0
Hydrangea serratifolia		5.3	1.9							0.7
Cissus striata						0				0

Table C.4: List of species found in each populations and its Midpoint coverage range (%) based on Braun Blanquet coverage abundance scale (Coversion of the Braun Blanquet to the Midpoint scale in Table C.5). Av. indicate the average of the coverage abundance across population by species. The species are ordered from highest to lowest percentage of coverage by population. sp refers to unidentified species

Plots Corral de Salas												
Species	**Life form**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Av. (%)**
Nothofagus obliqua	Tree	87.5	87.5	62.5								79.17
Prumnopitys andina	Tree	2.5	2.5	15								6.67
Aristotelia chilensis	Tree		0.1	0.1								0.07
Lomatia hirsuta	Tree			0.1								0.03
sp (Fabaceae)	Shrub	37.5		2.5								13.33
sp	Shrub	15										5.00
Chusquea sp	Herb	15	15	15								15.00
Plots Los Punquios												
Species	**Life form**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Av. (%)**
Prumnopitys andina	Tree	15	62.5	15	15	15	15	37.5	2.5	87.5	37.5	30.25
Nothofagus obliqua	Tree	2.5	2.5	37.5	2.5	37.5	37.5		62.5			18.25
Lomatia hirsuta	Tree	15		2.5		15	2.5		15	2.5		5.25
Luma apiculata	Tree	15		0.1							0.1	1.520
Austrocedrus chilensis	Tree	2.5	2.5		2.5	2.5	2.5					1.25

Table C.4 continued from previous page

Aristotelia chilensis	Tree	2.5				0.1				2.5		0.51
Laurelia sempervirens	Tree										2.5	0.25
Azara petiolaris	Shrub				15	37.5	15	0.1		15	62.5	14.51
Fabaceae sp	Shrub							15	2.5		37.5	5.50
Baccharis sp	Shrub					2.5						0.25
sp	Shrub					0.1		0.1	0.1			0.03
Chusquea sp	Herb	2.5		87.5		2.5					15	10.75
Bleschnum sp	Fern	62.5		2.5		15	2.5	15				9.75
Hydrangea serratifolia	Climbers	15			37.5							5.25

Plots Los Lleuques

Species	**Life form**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Av. (%)**
Prumnopitys andina	Tree	2.5	2.5	2.5	2.5	62.5	87.5	2.5	2.5	15		20.00
Nothofagus obliqua	Tree	15	2.5	15	15		2.5	15	37.5	15		13.06
Podocarpus salignus	Tree	15	37.5	15	15	2.5		2.5	2.5			10.00
Nothofagus dombeyi	Tree	15	37.5		15			2.5	2.5			8.06
Lomatia hirsuta	Tree	2.5	2.5	2.5		2.5		15	2.5	37.5		7.22
Dasyphyllum dicanthoides	Tree	37.5	2.5					15				6.11

Table C.4 continued from previous page

Drimys winteri	Tree	15	15	2.5	15			2.5	2.5			5.83
Embothrium coccineum	Tree			2.5		15		15				3.61
Luma apiculata	Tree	2.5	2.5									0.56
Gaultheria sp	Shrub					2.5	2.5			15		2.22
Azara serrata	Shrub		2.5	2.5								0.56
Azara microphylla	Shrub		2.5									0.28
Chusquea sp	Herb			15	2.5		2.5	2.5	2.5	15		4.00
Hydrangea serratifolia	Climbers	2.5	15									1.94
Baccharis sp									2.5			0.28
Colletia hystrix						2.5						0.28

Plots Laja

Species	**Life form**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Av. (%)**
Dasyphyllum dicanthoides	Tree	15	15	15	15	37.5	37.5	2.5	2.5	15	15	17.00
Prumnopitys andina	Tree	2.5	37.5	37.5	2.5	2.5	2.5	37.5	2.5	15	2.5	14.25
Lomatia dentata	Tree		15	37.5	37.5	15	2.5				2.5	11.00
Aextoxicon punctatum	Tree	2.5	2.5	15	15	15	2.5	15	2.5	2.5	15	8.75
Luma apiculata	Tree	2.5	2.5	15	15	15	2.5	15		2.5	2.5	7.25

Table C.4 continued from previous page

Nothofagus obliqua	Tree					2.5			2.5	15	15	3.50
Aristotelia chilensis	Tree	15	2.5				2.5	2.5	2.5	2.5		2.75
Nothofagus dombeyi	Tree		2.5		2.5				15			2.00
Lomatia hirsuta	Tree	2.5			2.5	2.5		2.5	2.5	2.5	2.5	1.75
Austrocedrus chilensis	Tree	2.5			2.5						2.5	0.75
Gevuina avellana	Tree	2.5				2.5			2.5			0.75
Maytenus boaria	Tree	2.5		2.5	2.5							0.75
Rhaphithamnus spinosus	Tree	2.5		2.5								0.50
Sophora microphylla	Tree									2.5	2.5	0.50
Azara petiolaris	Tree	2.5	0.1						0.1			0.27
Maytenus chilensis	Tree									2.5		0.250
Myrceugenia ovata	Tree		2.5									0.25
Schinus patagonicus	Tree									2.5		0.25
Berberis microphylla	Shrub	15										1.50
Gaultheria sp	Shrub	15										1.50
Azara microphylla	Shrub		2.5									0.25
Colletia sp	Shrub		2.5									0.25

Table C.4 continued from previous page

Myoschilo oblonga	Shrub		0.1									0.010
Heliotropium chilense	Herb								2.5			0.25
Plots Trapa-Trapa												
Species	**Life form**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Av. (%)**
Prumnopitys andina	Tree	87.5	87.5	87.5	87.5	87.5	87.5	87.5	87.5	87.5	62.5	85.00
Lomatia hirsuta	Tree			2.5		2.5			2.5		37.5	4.50
Nothofagus dombeyi	Tree									15	2.5	1.75
Nothofagus obliqua	Tree		2.5	2.5				2.5	2.5			1.00
Austrocedrus chilensis	Tree								0.1			0.01
Azara integrifolia	Shrub								0.1			0.01
Chusquea sp	Herb	2.5	2.5							15		2.00
Berberis sp									0.1			0.01
Plots Pua-Santa Lucia												
Species	**Life form**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Av. (%)**
Prumnopitys andina	Tree	37.5	62.5	15	15	37.5	87.5	37.5	87.5	62.5	87.5	53.00
Persea lingue	Tree			62.5	37.5		15	37.5	15	15	2.5	18.50
Nothofagus obliqua	Tree	15	15	15				15				6.00

Table C.4 continued from previous page

Rhaphithamnus spinosus	Tree	2.5	15			37.5						5.50
Lomatia dentata	Tree	2.5				37.5						4.00
Podocarpus salignus	Tree	2.5								15	2.5	2.00
Aristotelia chilensis	Tree			2.5	2.5							0.50
Luma apiculata	Tree							2.5				0.25
Rhamnus diffusus	Shrub	2.5		2.5	15							2.00
Rubus ulmifolius	Shrub	15	2.5									1.75
Chusquea sp	Herb	15	2.5		2.5	37.5						5.75
Cissus striata	Climbers	0.1										0.01

Plots Conguillio

Species	**Life form**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Av. (%)**
Prumnopitys andina	Tree	62.5	62.5	37.5	62.5	37.5	15	62.5	62.5	62.5	62.5	52.75
Nothofagus obliqua	Tree	15	15		2.5	37.5	37.5	2.5		2.5	15	12.75
Laurelia sempervirens	Tree			37.5	15							5.25
Gevuina avellana	Tree		2.5		2.5			2.5	15	15	2.5	4.00
Aristotelia chilensis	Tree					15	15	2.5			2.5	3.50
Nothofagus antartica	Tree					2.5	15		15			3.25

Table C.4 continued from previous page

Sophora microphylla	Tree	2.5			15	15						3.25
Lomatia hirsuta	Tree		2.5					2.5		15	2.5	2.25
Rhaphithamnus spinosus	Tree	2.5	2.5		2.5	2.5						1.00
Austrocedrus chilensis	Tree								2.5	2.5		0.50
Lomatia dentata	Tree	2.5								2.5		0.50
Ribes magellanicum	Shrub						15					1.50
Azara microphylla	Shrub							2.5				0.25

Plots Reigolil

Species	**Life form**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Av. (%)**
Prumnopitys andina	Tree	37.5	62.5	37.5	62.5	62.5	87.5	87.5	87.5	87.5	87.5	70.00
Nothofagus obliqua	Tree	37.5		37.5	15	15	15	15	15			15.00
Lomatia hirsuta	Tree	37.5		2.5		2.5				2.5		4.50
Nothofagus antartica	Tree	2.5	2.5								15	2.00
Rhaphithamnus spinosus	Tree	2.5		2.5								0.50
Myrceugenia ovata	Tree			2.5								0.25
Nothofagus dombeyi	Tree		2.5									0.25
Azara microphylla	Shrub			2.5	2.5							0.50

Table C.4 continued from previous page

Chusquea sp	Herb				2.5							0.25
		Plots Molulco										
Species	**Life form**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Av.(%)**
Prumnopitys andina	Tree	87.5	62.5	87.5	87.5	87.5	87.5	87.5	87.5	87.5	87.5	85.00
Nothofagus dombeyi	Tree		15									1.50
Nothofagus obliqua	Tree			2.5		2.5						0.50
Aristotelia chilensis	Tree			0.1								0.01

Table C.5: Conversion from Braun Blanquet scale to Midpoint of coverage range (%)

Braun-Blanquet scale	Rage of cover (%)	Midpoint of cover range (%)
5	75-100	87.5
4	50-75	62.5
3	25-50	37.5
2	May-25	15
1	<5 (numerous individuals)	2.5
+	<5 (few individuals)	0.1

Table C.6: Braun Blanquet cover abundance in each population. sp refers to unidentified species.

Plots Corral de Salas											
Species	**Life form**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Prumnopitys andina	Tree	1	1	2							
Aristotelia chilensis	Tree		+	+							
Lomatia hirsuta	Tree			+							
Nothofagus obliqua	Tree	5	5	4							
sp (Fabaceae)	Shrub	3		1							
Chusquea sp	Herb	2	2	2							
sp		2									
Plots Los Punquios											
Species	**Life form**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Prumnopitys andina	Tree	2	4	2	2	2	2	3	1	5	3
Aristotelia chilensis	Tree	1				+				1	
Austrocedrus chilensis	Tree	1	1		1	1	1				
Laurelia semperverins	Tree										1
Lomatia hirsuta	Tree	2		1		2	1		2	1	
Luma apiculata	Tree	2		+							+
Nothofagus obliqua	Tree	1	1	3	1	3	3		4		
Azara microphylla	Shrub										
Azara petiolaris	Shrub				2	3	2	+		2	4
Baccharis sp	Shrub					1					
Fabaceae sp	Shrub							2	1		3

Table C.6 continued from previous page

sp	Shrub					+		+	+		
Chusquea sp	Herb	1		5		1					2
Bleschnum sp	Fern	4		1		2	1	2			
Hydrangea serratifolia	Climbers	2			3						

Plots Los Lleuques

Species	**Life form**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Prumnopitys andina	Tree	1	1	1	1	4	5	1	1	2	
Dasyphyllum dicanthoides	Tree	3	1					2			
Drimys winteri	Tree	2	2	1	2			1	1		
Embothrium coccineum	Tree			1		2		2			
Lomatia hirsuta	Tree	1	1	1		1		2	1	3	
Luma apiculata	Tree	1	1								
Nothofagus dombeyi	Tree	2	3		2			1	1		
Nothofagus obliqua	Tree	2	1	2	2		1	2	3	2	
Podocarpus salignus	Tree	2	3	2	2	1		1	1		
Azara microphylla	Shrub		1								
Azara serrata	Shrub		1	1							
Gaultheria sp	Shrub					1	1			2	
Chusquea sp	Herb			2	1		1	1	1	2	
Hydrangea serratifolia	Climbers	1	2								
Baccharis sp									1		
Colletia hystrix						1					

Plots Laja

Species	**Life form**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Prumnopitys andina	Tree	1	3	3	1	1	1	3	1	2	1
Aextoxicon punctatum	Tree	1	1	2	2	2	1	2	1	1	2
Aristotelia chilensis	Tree	2	1				1	1	1	1	
Austrocedrus chilensis	Tree	1			1						1
Azara petiolaris	Tree	1	+						+		

Table C.6 continued from previous page

Dasyphyllum dicanthoides	Tree	2	2	2	2	3	3	1	1	2	2
Gevuina avellana	Tree	1				1			1		
Lomatia dentata	Tree		2	3	3	2	1				1
Lomatia hirsuta	Tree	1			1	1		1	1	1	1
Luma apiculata	Tree	1	1	2	2	2	1	2		1	1
Maytenus boaria	Tree	1		1	1						
Maytenus chilensis	Tree									1	
Myrceugenia ovata	Tree		1								
Nothofagus dombeyi	Tree		1		1				2		
Nothofagus obliqua	Tree					1			1	2	2
Rhaphithamnus spinosus	Tree	1		1							
Schinus patagonicus	Tree									1	
Sophora microphylla	Tree									1	1
Azara microphylla	Shrub		1								
Berberis microphylla	Shrub	2									
Colletia sp	Shrub		1								
Gaultheria sp	Shrub	2									
Myoschilo oblonga	Shrub		+								
Heliotropium chilense	Herb								1		

Plots Trapa-Trapa

Species	**Life form**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Prumnopitys andina	Tree	5	5	5	5	5	5	5	5	5	4
Austrocedrus chilensis	Tree								+		
Lomatia hirsuta	Tree			1		1			1		3
Nothofagus dombeyi	Tree									2	1
Nothofagus obliqua	Tree		1	1				1	1		
Azara integrifolia	Shrub								+		
Chusquea sp	Herb	1	1							2	
Berberis sp									+		

Table C.6 continued from previous page

Plots Pua-Santa Lucia											
Species	**Life form**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Prumnopitys andina	Tree	3	4	2	2	3	5	3	5	4	5
Aristotelia chilensis	Tree			1	1						
Lomatia dentata	Tree	1				3					
Luma apiculata	Tree							1			
Nothofagus obliqua	Tree	2	2	2				2			
Persea lingue	Tree			4	3		2	3	2	2	1
Podocarpus salignus	Tree	1								2	1
Rhaphithamnus spinosus	Tree	1	2			3					
Rhamnus diffusus	Shrub	1		1	2						
Rubus ulmifolius	Shrub	2	1								
Chusquea sp	Herb	2	1		1	3					
Cissus striata	Climbers	+									
Plots Conguillio											
Species	**Life form**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Prumnopitys andina	Tree	4	4	3	4	3	2	4	4	4	4
Aristotelia chilensis	Tree					2	2	1			1
Austrocedrus chilensis	Tree								1	1	
Gevuina avellana	Tree		1		1			1	2	2	1
Laurelia semperverins	Tree			3	2						
Lomatia dentata	Tree	1								1	
Lomatia hirsuta	Tree		1					1		2	1
Nothofagus antartica	Tree					1	2		2		
Nothofagus obliqua	Tree	2	2		1	3	3	1		1	2
Rhaphithamnus spinosus	Tree	1	1		1	1					
Sophora microphylla	Tree	1			2	2					
Azara microphylla	Shrub							1			
Ribes magellanicum	Shrub						2				

Table C.6 continued from previous page

Plots Reigolil											
Species	**Life form**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Prumnopitys andina	Tree	3	4	3	4	4	5	5	5	5	5
Lomatia hirsuta	Tree	3		1		1				1	
Myrceugenia ovata	Tree			1							
Nothofagus antartica	Tree	1	1								2
Nothofagus dombeyi	Tree		1								
Nothofagus obliqua	Tree	3		3	2	2	2	2	2		
Rhaphithamnus spinosus	Tree	1		1							
Azara microphylla	Shrub			1	1						
Chusquea sp	Herb				1						
Plots Molulco											
Species	**Life form**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Prumnopitys andina	Tree	5	4	5	5	5	5	5	5	5	5
Aristotelia chilensis	Tree			+							
Nothofagus dombeyi	Tree		2								
Nothofagus obliqua	Tree			1		1					

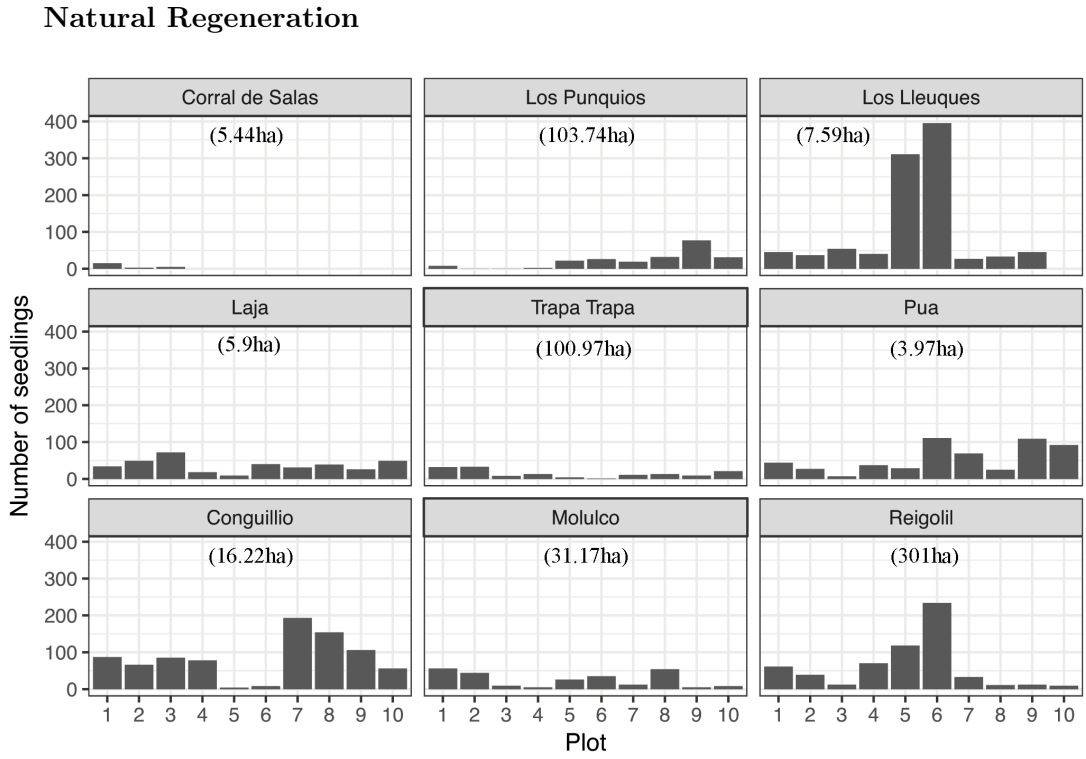


Figure C.1: Distribution of seedlings (*Pr. andina* and other species combined) for each plot in each population. Populations areas (ha) are included. Corral de Salas was surveyed with three plots and los Punquios with nine populations.

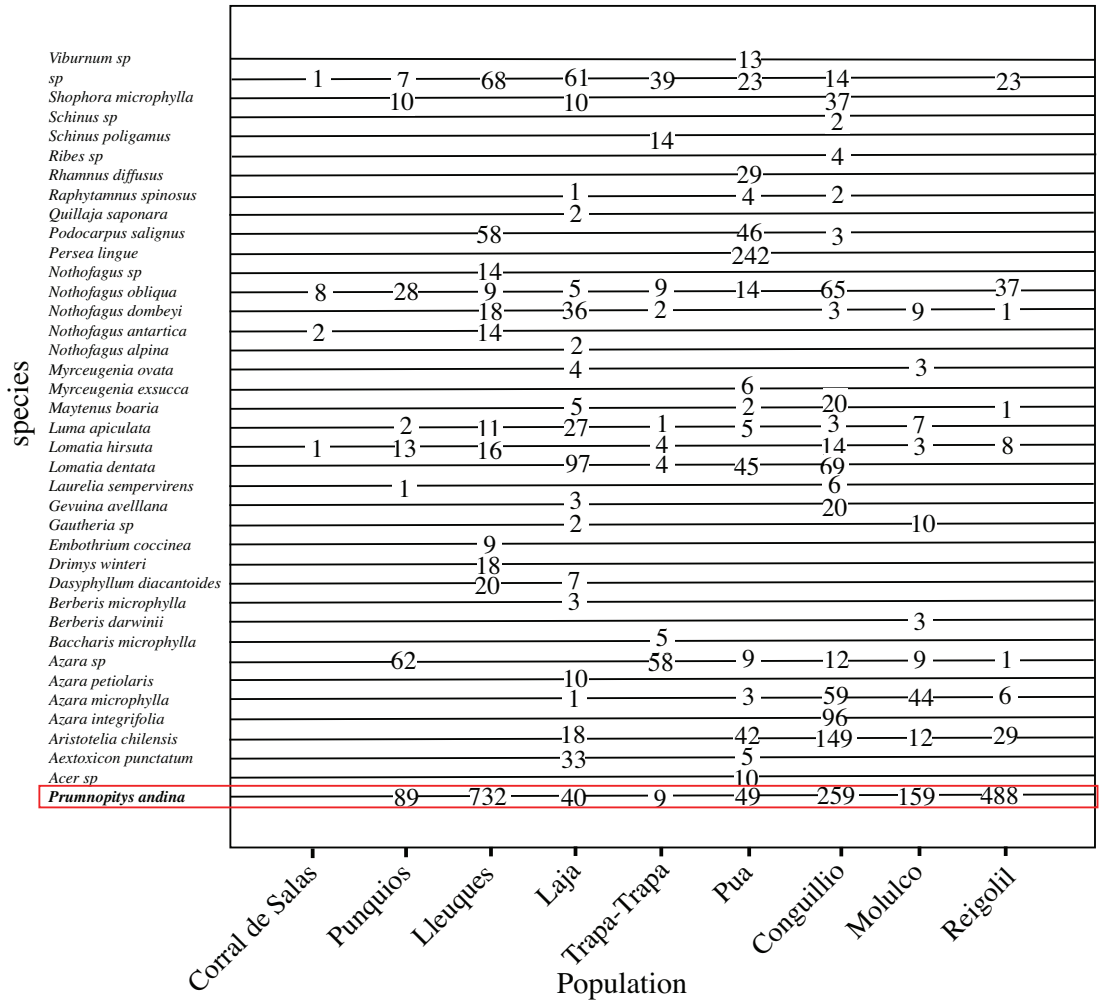


Figure C.2: Total number of seedlings by species found in each population. Populations are listed from the north (left) to the south (right). Species are listed alphabetically, except for *Pr. andina*.

Statistical analyses

Table C.7: Test of normality for factors potentially associated with variation in regeneration.

	Kolmogorov-Smirnova			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Regeneration *P. andina*	0.355	82	0.000	0.385	82	0.000
Regeneration other sp	0.159	82	0.000	0.831	82	0.000
Male	0.303	82	0.000	0.657	82	0.000
Female	0.324	82	0.000	0.66	82	0.000
Total adults	0.169	82	0.000	0.877	82	0.000
Total regeneration	0.235	82	0.000	0.622	82	0.000
Mean light intensity	0.448	82	0.000	0.461	82	0.000

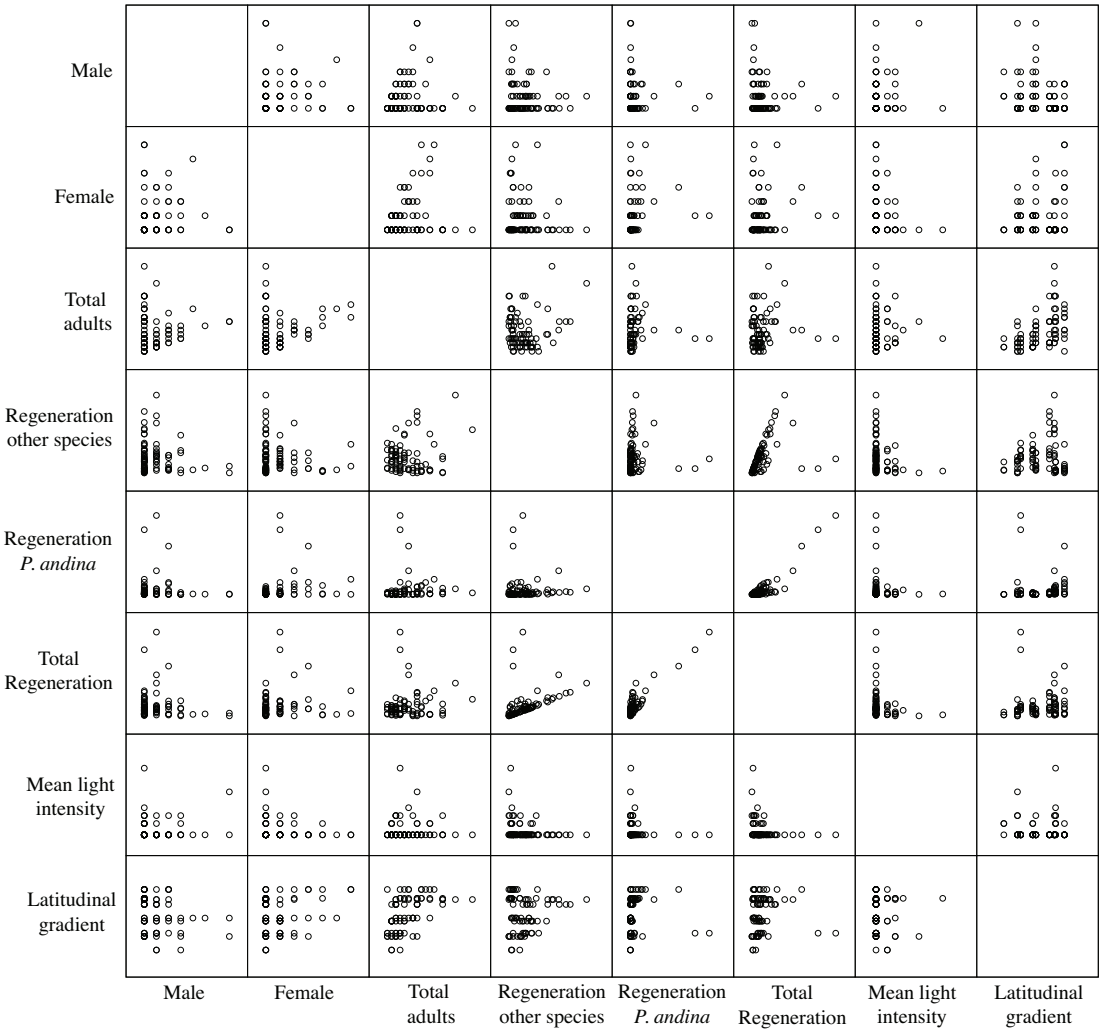


Figure C.3: Matrix-scatter plot showing the data distribution for factors potentially associated with variation in regeneration.

Appendix D

Appendix from Chapter 5

Threats found in each conifer species population

This section presents the threats observed throughout the populations surveyed for the four study species, including their level of impact.

The threats were recorded *in-situ* in ten populations of each conifer species (except in *F. cupressoides* where only nine populations were included), covering their entire natural ranges. The threat classification was assessed following the hierarchical structure of threat types provided by the IUCN Red List. This list includes direct or indirect threats by human activities or processes that have impacted or could impact the status of the species being assessed (Full details in Threats Classification Scheme Version 3.2: in [1]). The full threat list proposed by the IUCN was reduced according to the potential pressures likely to be found in the Temperate Chilean rain forest (List in Table D.1).

For each threat, I also recorded the timing of the threat (i.e. past, ongoing or future), its scope (i.e. the proportion of the total population affected) and severity (i.e the overall declines caused by the threat). This was used to evaluate the impact of each threat (details in Table D.2 and Figure D.1).

[1] https://www.iucnredlist.org/resources/threat-classification-scheme

Table D.1: List of Threats Classification Scheme Version 3.2. Here I list only the pressures likely to be found in the Chilean conifer forests

Residential and commercial development
Housing and urban areas
Commercial and industrial areas
Tourism and recreation areas
Agriculture
Annual and perennial non-timber crops
Wood and Pulp plantations
Livestock farming
Energy production and mining
Renewable energy
Transportation and service corridors
Roads and railroads
Utility and service line
Biological resource use
Logging and wood harvesting
Natural system modifications
Fire and fires supression
Dams and water management/use
Other ecosystmen modification
Invasive and other problematic species, gene and diseases
Invasive non-native/alien species / diseases
Problematic native species/diseases
Introduced genetic material
Problematic species/diseases of unknown origin
Geological events
Volcanoes
Earthquakes/tsumanis
Avalanches/landslides
Droughts
Temperature extremes
Storms and flooting
Other impacts
Other threat

Table D.2: Score system provide by the IUCN. Timing, scope and severity option of each threat and their respectively score provided by the IUCN.

Score	Timing option
4	On going
3	In the past but now suspended and likely to return
2	Only in the future
1	Only in the past and unlikely to return
0	Unknown
Score	**Scope option**
3	Affects the whole population more than 90%
2	Affects the majority of the population 50-90%
1	Affect the minority of the population less than 50%
0	Unknown
Score	**Severity option**
3	Very rapid decline (>30% over 10 years or three generations; whichever is the longer)
2	Rapid decline (20–30% over 10 years or three generations; whichever is the longer)
1	Slow (<20% over 10 years or three generations; whichever is the longer)
1	Fluctuating (Causing or likely to cause fluctuations)
0	Negligible (Causing or likely to cause negligible declines)
0	No decline
0	Unknown

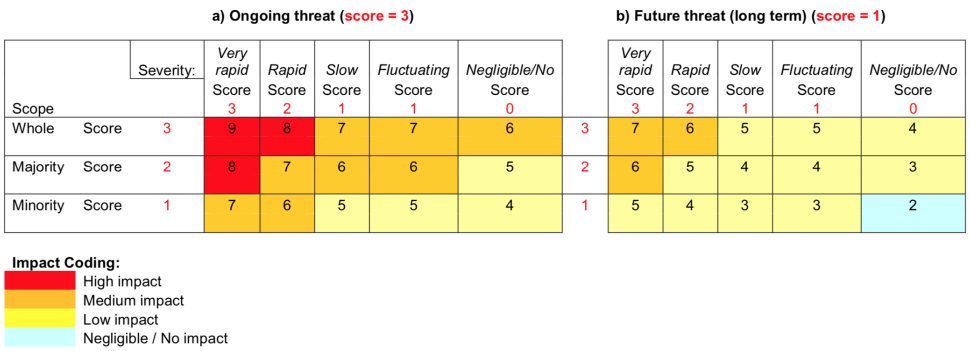


Figure D.1: Level of impact conversion based on the score system provide by the IUCN. The table was extracted from the IUCN web-site; https://www.iucnredlist.org/resources/threat-classification-scheme

Results

Threats found by species

The most common threats found in all conifer species were associated with residential and commercial development, agriculture, biological resource use and geological events.

Saxegothaea conspicua

1. **Residential and commercial development**
    - Housing and urban areas, was observed in 20% of the populations. The threat was categorised as a low-impact threat in all populations (details Table D.4).
    - Tourism and recreation areas, was observed in 70% of the populations. The threat was categorised as a low-impact threat in all populations.
2. **Agriculture**
    - Wood and Pulp plantations, was observed in 30% of the populations. The threat was categorised as a medium-impact threat in all populations.
    - Livestock farming, was observed in 30% of the populations. The threat was categorised as a medium-impact threat in all populations.
3. **Biological resource use**
    - Logging and wood harvesting, was observed in 30% of the populations. The threat was categorised as a low-impact threat in all populations.
    - **Geological events**
    - Drought, was observed in 10% of the populations. The threat was categorised as a medium-impact threat in all populations.

Prumnopitys andina

1. **Residential and commercial development**

    - Housing and urban areas (threat category), was observed in 40% of the populations with a low-impact level (details Table D.4).
    - Tourism and recreation areas were observed in 30% of the populations. Low-impact in all populations.

2. **Agriculture**

    - Wood and Pulp plantations were observed in 30% of the populations. The threat was categorised as a medium and high-impact threat, depending on the population.
    - Livestock farming was observed in 100% of the populations. The threat was categorised as a medium-impact threat in all populations.

3. **Biological resource use**

    - Logging and wood harvesting was observed in 40% of the populations. The threat was categorised as a low and medium-impact threat depending on the population.

Podocarpus salignus

1. **Residential and commercial development**

    - Housing and urban areas, was observed in 50% of the populations. The threat was categorised as a low-impact threat in all populations (details Table D.5).
    - Tourism and recreation areas, was observed in 40% of the populations. The threat was usually categorised as a low-impact threat in most populations (in only one sigle population this threat was categorised as a high impact threat).

2. **Agriculture**

    - Wood and Pulp plantations, was observed in 50% of the populations. The threat was categorised as a medium and high-impact threat in all populations.
    - Livestock farming, was observed in 40% of the populations. The threat was categorised as a medium and high-impact threat in all populations.

3. **Biological resource use**

    - Logging and wood harvesting, was observed in 30% of the populations. The threat was categorised as a low-impact threat in all populations.

    - **Geological events**
    - Drought, was observed in 30% of the populations. The threat was categorised as a medium and high-impact threat.

Fitzroya cupressoides

1. **Residential and commercial development**

    - Housing and urban areas, was observed in 30% of the populations. The threat was categorised as a low-impact threat in all populations (details Table D.5).
    - Tourism and recreation areas, was observed in 40% of the populations. The threat was categorised as a low-impact threat in all populations.

2. **Agriculture**

    - Wood and Pulp plantations, was observed in 20% of the populations. The threat was categorised as a medium and high-impact threat.
    - Livestock farming, was observed in 20% of the populations. The threat was categorised as a low and medium-impact threat.

Table D.3: List of the main threats found in each of the species populations and its percentage of occurrence across the entire natural distribution of the species.

		Percentage (%) of occurrence across populations			
Ongoing threat types by category	**Threat**	***S. conspicua***	***Pr. andina***	***P. salignus***	***F. cupressoides***
Residential and	Housing and urban areas	20	40	50	30
commercial development	Tourism and recreation areas	70	30	40	40
Agriculture	Wood and Pulp plantations	30	30	50	20
	Livestock farming	30	100	40	20
Biological resource use	Logging and wood harvesting	30	40	30	
Geological events	Drought	10		30	

Table D.4: List of threats and the level of impact on the *S. conspicua* and *Pr. andina* species population. Here are listed only threats that were observed at least in one population. Populations are listed with numbers (1-10) from the northern (1) to the southern (10) population for each species. The number 1 in each box (filled by colour) means the presence of the threat in that specific population. The colours indicate the level of impact of the threat in that specific population. Red indicate a hight impact, orange indicate a medium impact and yellow indicate a low impact according to the score system provided by the IUCN.

Type of threat	***S. conspicua***										***Pr. andina***									
	populations										populations									
Residential and commercial development	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
Housing and urban areas									1	1		1	1		1					1
Tourism and recreation areas	1		1	1	1		1		1	1	1	1						1		
Agriculture																				
Wood and Pulp plantations	1	1							1					1	1	1				
Livestock farming		1				1			1		1	1	1	1	1	1	1	1	1	1
Transportation and service corridors																				
Roads and railroads								1				1								
Biological resource use																				
Logging and wood harvesting		1							1	1			1	1	1					1
Natural system modifications																				
Dams and water management/use												1								

Table D.4 continued from previous page

Invasive non-native/alien species / diseases									
Problematic native species/diseases					1				
Geological events									
Volcanoes				1			1		
Avalanches/landslides									
Droughts		1							

Population *S. conspicua*: (1)Nahuelbuta, (2)Villas las Araucarias, (3)Huerquehue, (4)Villarica, (5)Oncol, (6)Llancahue, (7)Huilo Huilo,(8)Huinay, (9) Rio Cisne (10)Rio Futa.

Population *Pr. andina*: (1)Corral de Salas, (2)Los Punquios, (3)Los Lleuques, (4)Laja, (5)Trapa-Trapa, (6)Nahuelbuta, (7)Pua ,(8)Conguillio, (9)Molulco, (10)Reigolil.

Table D.5: List of threats and the level of impact on the *P. salignus* and *F. cupressoides* species population. Here are listed only threats that were observed at least in one population. Populations are listed with numbers (1-10 for *P. salignus* and 1-9 for *F. cupressoides*) from the northern (1) to the southern (9 or 10) population for each conifer species. The number 1 in each box (filled by colours) represent the presence of the threat in that specific population. The colours indicate the level of impact of the threat in that specific population. Red indicate a hight impact, orange indicate a medium impact and yellow indicate a low impact according to the score system provided by the IUCN.

Type of threat	***P. salignus***										***F. cupressoides***								
	populations										populations								
Residential and commercial development	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9
Housing and urban areas		1	1	1		1				1				1		1			1
Tourism and recreation areas		1		1			1	1				1	1				1		1
Agriculture																			
Annual and perennial non-timber crops								1											
Wood and Pulp plantations					1	1	1	1		1	1					1			
Livestock farming							1	1	1	1			1			1			
Energy production and mining																			
Renewable energy	1																		
Transportation and service corridors																			
Roads and railroads	1																	1	
Biological resource use																			

Table D.5 continued from previous page

Logging and wood harvesting				1		1		1		
Natural system modifications										
Dams and water management/use	1									
Other ecosystmen modification	1	1								
Invasive non-native/alien species / diseases										
Problematic species/diseases of unknown origin						1				
Geological events										
Volcanoes									1	1
Earthquakes/tsumanis										
Avalanches/landslides										
Droughts				1		1		1		

Population *P. salignus*: (1)Hornillos, (2)Los Lleuques, (3)Antuco, (4)Nahuelbuta, (5)Los Guindos , (6)Santa Lucia , (7)Tirua,(8)Llanacahue, (9)Reumen (10)Llancacura.

Population *F. cupressoides*: (1)ANCHILE, (2)Alerce Costero, (3)Cordillera Pelada, (4)Rio Pescado, (5)Fundo Nuñez, (6)Astillero, (7)Alerce andino ,(8)Huinay, (9)Caleta Gonzalo.